

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 15:06:47 ; Search time 3759 Seconds
(without alignments)
2688.822 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402
Sequence: 1 ccggcgccggcgtccacagct.....ggcatactctgtgtcc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapect 60.0

Searched: 24791104 seqs, 12571243825 residues

Word size : 0

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

1: /cgn2_6/ptodata/1/pna/US0600.COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US0601.COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US0602.COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US0603.COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US0604.COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US0605.COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US0606.COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US0607.COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US0608.COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US0609.COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US0610.COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US0611.COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US0612.COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US0613.COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US0614.COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US0615.COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US0616.COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US0617.COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US0618.COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US0619.COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US0620.COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US0621.COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US0622.COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US0623.COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US0624.COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US0625.COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US0626.COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US0627.COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US0628.COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US0629.COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US0630.COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US0631.COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US0632.COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US0633.COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US0634.COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US0635.COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US0636.COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US0637.COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US0638.COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US0639.COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US0640.COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US0641.COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US0642.COMB.seq:*

Result				Query		ID		Description
No.	Score	Match	Length	DB				
1	402	100.0	402	25	US-09-654-617-437554			Sequence 457554, A
2	402	100.0	402	27	US-09-684-016-457554			Sequence 457554, A
3	402	100.0	402	32	US-09-850-147-1			Sequence 1, Appl
4	402	100.0	402	64	US-60-202-213-1			Sequence 1, Appl
5	29	7.2	341	33	US-09-865-439A-8889			Sequence 8889, Ap
6	29	7.2	351	33	US-09-865-439A-8361			Sequence 4361, Ap
7	29	7.2	385	32	US-09-850-147-10			Sequence 10, Appl
8	29	7.2	385	64	US-09-865-439A-18142			Sequence 18142, A
9	29	7.2	393	22	US-09-565-306-82924			Sequence 82924, A
10	29	7.2	426	32	US-09-850-147-40			Sequence 40, Appl
11	29	7.2	426	64	US-60-202-213-40			Sequence 40, Appl
12	29	7.2	433	32	US-09-850-147-1092			Sequence 1092, Ap
13	29	7.2	433	64	US-60-202-213-1089			Sequence 1089, Ap
14	29	7.2	450	64	US-60-207-458-54563			Sequence 54563, Ap
15	29	7.2	456	25	US-09-654-617-451847			Sequence 451847, A
16	29	7.2	456	27	US-09-684-016-451847			Sequence 15731, A
17	29	7.2	484	33	US-09-865-419A-15731			Sequence 2535, Ap
18	29	7.2	484	64	US-60-208-063-2535			Sequence 41873, A
19	29	7.2	545	33	US-09-873-402A-41873			Sequence 51912, A
20	29	7.2	550	33	US-09-865-419A-51912			
21	29	7.2	550	33	US-09-865-419A-51912			

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

C	22	29	7.2	559	33	US-09-873-402A-546567	Sequence 546567, A
C	23	29	7.2	559	64	US-60-209-830-A46016	Sequence 46016, A
C	24	29	7.2	570	33	US-09-865-4339A-116358	Sequence 116358, A
C	25	29	7.2	605	33	US-09-865-4339A-93607	Sequence 93607, A
C	26	29	7.2	605	64	US-60-207-458-137622	Sequence 137622, A
C	27	29	7.2	660	33	US-09-861-478-52	Sequence 52, App
C	28	29	7.2	660	33	US-09-861-478-506	Sequence 506, App
C	29	29	7.2	995	42	US-10-219-999-1500	Sequence 1500, App
C	30	29	7.2	1231	42	US-10-219-999-20024	Sequence 20024, A
C	31	29	7.2	1325	42	US-10-219-999-18314	Sequence 18314, A
C	32	28	7.0	1694	25	US-09-654-617-262628	Sequence 262628, A
C	33	28	7.0	1694	27	US-09-684-016-262628	Sequence 262628, A
C	34	27	6.7	345	64	US-60-207-458-47364	Sequence 47364, A
C	35	27	6.0	397	64	US-60-207-458-43952	Sequence 43952, A
C	36	22	5.5	262	16	US-09-244-000A-14941	Sequence 14941, A
C	37	22	5.5	262	36	US-09-878-703-14941	Sequence 14941, A
C	38	22	5.5	349	19	US-09-521-640-256139	Sequence 256139, A
C	39	22	5.5	349	58	US-60-140-769-43520	Sequence 43520, A
C	40	22	5.5	396	19	US-09-521-640-144941	Sequence 144941, A
C	41	22	5.5	396	19	US-09-521-640-156054	Sequence 156054, A
C	42	22	5.5	411	33	US-09-865-419A-99944	Sequence 9944, App
C	43	22	5.5	574	24	US-09-620-246-335	Sequence 335, App
C	44	22	5.5	574	39	US-10-062-234-335	Sequence 335, App
C	45	22	5.5	607	33	US-09-874-708A-77966	Sequence 77966, A

ALIGNMENTS

```

RESULT 1
US-09-654-617-457554
: Sequence 457554, Application US/09654617
: GENERAL INFORMATION:
: APPLICANT: Koyalic, David K.
: APPLICANT: Liu, Jingdong
: TITLE OF INVENTION: Annotated Plant Genes
: FILE REFERENCE: 38-21(15097)D
: CURRENT APPLICATION NUMBER: US/09/654,617
: CURRENT FILING DATE: 2000-09-05
: NUMBER OF SEQ ID NOS: 463173
: SEQ ID NO 457554
: LENGTH: 402
: TYPE: DNA
: ORGANISM: Sorghum bicolor
: US-09-654-617-457554

```

Query Match	100.0%;	Score 402;	DB 25;	Length 402;
-------------	---------	------------	--------	-------------

Best Local Similarity 100.0%; Pred. No. 2.9e-212;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Sequence 54567, A
Sequence 46016, A
Sequence 116358, A
Sequence 93607, A
Sequence 1376722, A
Sequence 52, App
Sequence 506, App
Sequence 1500, App
Sequence 20024, A
Sequence 18314, A
Sequence 262628, A
Sequence 262628, A
Sequence 47364, A
Sequence 43952, A
Sequence 43952, A
Sequence 14911, A
Sequence 14911, A
Sequence 256139, A
Sequence 43520, A
Sequence 144941, A
Sequence 160504, A
Sequence 9944, App
Sequence 335, App
Sequence 335, App
Sequence 77966, A

```

QY      361 AGTACAAAACCAACGAGGAGGCGCATACCTCGTGTGCC 4020
        |||||||
        |||||||
        |||||||
Db       361 AGTACAAAACCAACGAGGAGGCGCATACCTCGTGTGCC 4020

RESULT 2
US-09-684-016-457554
; Sequence 457554, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovallc, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684.016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIORITY FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 457554
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
US-09-684-016-457554

```

RESULT 2

```

US-09-684-016-457554
; Sequence 457554, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingsong
; TITLE OF INVENTION: Annotated Plant Gen
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 457554
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
US-09-684-016-457554

```

Query Match	100.0%	Score 402;	DB 27;	Length 402;
Best Local Similarity	100.0%;	Pred. No. 2.9e-212;		
Matches 402; Conservative	0;	Mismatches	0;	Gaps 0;

QY	1	CCGGCCCGGGCTCACCACCTGGCCCACTGTAAGGCTTTAAAGACATGTGCCAGAGAAGGG	60
Db	1	CCGGCCCGGGCTCACACACTGGCCCACTGAAGGCTTTAAAGACATGTGCCAGAGAAGGG	60
QY		CCTTCACCTTGTGTGTGTGCACACCTAATTTATTTGCTCAGATGTGGCAATGAGAGCTTCAT	120
Db		CCTTCACCTTGTGTGTGTGCACACCTAATTTATTTGCTCAGATGTGGCAATGAGAGCTTCAT	120
QY	121	ATTAACTTCGATGGAAGAAAGGAAGAGATGTCAAGGCTTTCACAGAGCCGAGAAA	180
Db	121	ATTAACTTCGATGGAAGAAAGGAAGAGATGTCAAGGCTTTCACAGAGCCGAGAAA	180
QY	181	ACAACCAATTCOCAGGGCCAAAGGCGTTGGAAATCCATATTCCTTGAATTCAGAGGAC	240
Db	181	ACAACCAATTCOCAGGGCCAAAGGCGTTGGAAATCCATATTCCTTGAATTCAGAGGAC	240
QY	241	AATTTAGAAAGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGGTGTAAAT	300
Db	241	AATTTAGAAAGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGGTGTAAAT	300
QY	301	GTGAGAGTTCAGATCATATGCAATCTGATGATTCATTTTGGTTCTTTTGAACACA	360
Db	301	GTGAGAGTTCAGATCATATGCAATCTGATGATTCATTTTGGTTCTTTTGAACACA	360

```

QY      301 GGGAGGTTCAATCATTAAGCCATCGATGCATTCTTGTTTGAACACA   360
        |||
Db       301 GGAGAGGTTCAATCATTAAGCCATCGATGCATTCTTGTTTGAACACA   360
        |||

QY      361 AGTACAAAAAACCCAGCGGAGGGGCATTAAGTCCTGTGTCC    402
        |||
Db       361 AGTACAAAAAACCCAGCGGAGGGGCATTAAGTCCTGTGTTC    402
        |||

RESULT 3
US-09-850-147-1
; Sequence 1, Application US-/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
;
```

;; PRIOR APPLICATION NUMBER: US 09/684,016
;; PRIOR FILING DATE: 2000-10-10
;; NUMBER OF SEQ ID NOS: 18014
;; SEQ ID NO 1
;; LENGTH: 402
;; TYPE: DNA
;; ORGANISM: Sorghum bicolor
;; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-09-850-147-1

Query Match 100.0%; Score 402; DB 32; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.9e-212;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGCCGGGCTCACCAGCTGGGCCAGTGAAGGCTTAAAGACATGCGCAGAGAGAGGG 60
DB 1 CCGGGCCGGGCTCACCAGCTGGGCCAGTGAAGGCTTAAAGACATGCGCAGAGAGAGGG 60
QY 61 CCGTGAAGCTTGT 120
DB 61 CCGTGAAGCTTGT 120
QY 121 ATTAAGCTTCGATGGAAGAGAGAGAGATGCAAGGCTTCCACAGACCGAGAA 180
DB 121 ATTAAGCTTCGATGGAAGAGAGAGAGATGCAAGGCTTCCACAGACCGAGAA 180
QY 181 ACAACCAATTCAGAGGCGCAAGGGCTTGAATCCATATTCCTTGAAGGAGGAC 240
DB 181 ACAACCAATTCAGAGGCGCAAGGGCTTGAATCCATATTCCTTGAAGGAGGAC 240
QY 241 AATTTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGTTAAAT 300
DB 241 AATTTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGTTAAAT 300
QY 301 GTGGAGGTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 GTGGAGGTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AGTACAAAAAACCAACGGAGGGGAGGATTAACCTCTGTGTTC 402
DB 361 AGTACAAAAAACCAACGGAGGGGAGGATTAACCTCTGTGTTC 402

RESULT 4
US-60-202-213-1

;; Sequence 1, Application US/60202213
;; GENERAL INFORMATION:
;; APPLICANT: Andersen, Scott E.
;; APPLICANT: Edgerton, Michael D
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
;; FILE REFERENCE: 38-21(51914)A
;; CURRENT APPLICATION NUMBER: US/60/202,213
;; CURRENT FILING DATE: 2000-05-08
;; NUMBER OF SEQ ID NOS: 17986
;; SEQ ID NO 1
;; LENGTH: 402
;; TYPE: DNA
;; ORGANISM: Sorghum bicolor
;; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-60-202-213-1

Query Match 100.0%; Score 402; DB 64; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.9e-212;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGCCGGGCTCACCAGCTGGGCCAGTGAAGGCTTAAAGACATGCGCAGAGAGAGGG 60
DB 1 CCGGGCCGGGCTCACCAGCTGGGCCAGTGAAGGCTTAAAGACATGCGCAGAGAGAGGG 60
QY 61 CCGTGAAGCTTGT 120
DB 61 CCGTGAAGCTTGT 120

QY 121 ATTAAGCTTCGATGGAAGAGAGAGATGCAAGGCTTCCACAGACCGAGAA 180
DB 121 ATTAAGCTTCGATGGAAGAGAGAGATGCAAGGCTTCCACAGACCGAGAA 180
QY 181 ACAACCAATTCAGAGGCGCAAGGGCTTGAATCCATATTCCTTGAAGGAGTGCAC 240
DB 181 ACAACCAATTCAGAGGCGCAAGGGCTTGAATCCATATTCCTTGAAGGAGTGCAC 240
QY 241 AATTTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGTTAAAT 300
DB 241 AATTTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGTTAAAT 300
QY 301 GTGGAGGTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 GTGGAGGTTCAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AGTACAAAAAACCAACGGAGGGGAGGATTAACCTCTGTGTTC 402
DB 361 AGTACAAAAAACCAACGGAGGGGAGGATTAACCTCTGTGTTC 402

RESULT 5
US-09-865-439A-8889/C

;; Sequence 8889, Application US/09865439A
;; GENERAL INFORMATION:
;; APPLICANT: Edgerton, Michael D
;; APPLICANT: Hardeman, Kristine J.
;; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;; FILE REFERENCE: 38-21(51936)B
;; CURRENT APPLICATION NUMBER: US/09/865,439A
;; CURRENT FILING DATE: 2001-05-29
;; PRIOR APPLICATION NUMBER: US 60/207,458
;; PRIOR FILING DATE: 2000-05-30
;; NUMBER OF SEQ ID NOS: 119126
;; SEQ ID NO 8889
;; LENGTH: 341
;; TYPE: DNA
;; ORGANISM: Zea mays
;; FEATURE:
;; OTHER INFORMATION: Clone ID: LIB3205-106-P1-N1-B3
US-09-865-439A-8889

Query Match 7.2%; Score 29; DB 33; Length 341;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTTCTATATTAAGCTTCGATGAGAGA 140
DB 336 AGCTTCTATATTAAGCTTCGATGAGAGA 308

RESULT 6
US-09-865-439A-4361/C

;; Sequence 4361, Application US/09865439A
;; GENERAL INFORMATION:
;; APPLICANT: Edgerton, Michael D
;; APPLICANT: Hardeman, Kristine J.
;; APPLICANT: La Rosa, Thomas J.
;; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
;; FILE REFERENCE: 38-21(51936)B
;; CURRENT APPLICATION NUMBER: US/09/865,439A
;; CURRENT FILING DATE: 2001-05-29
;; PRIOR APPLICATION NUMBER: US 60/207,458
;; PRIOR FILING DATE: 2000-05-30
;; NUMBER OF SEQ ID NOS: 119126
;; SEQ ID NO 4361
;; LENGTH: 351
;; TYPE: DNA
;; ORGANISM: Zea mays

```
FEATURE:
; OTHER INFORMATION: Clone ID: LIB3205-049-P1-N1-F9
US-09-865-439A-4361

Query Match
Best Local Similarity 100.0%; Score 29; DB 33; Length 351;
Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTCTATATTAAGCTTCGATGAGAGA 140
DB 277 AGCTCTATATTAAGCTTCGATGAGAGA 249

RESULT 7
US-09-850-147-10
; Sequence 10, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 10
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-09-850-147-10

Query Match
Best Local Similarity 100.0%; Score 29; DB 33; Length 385;
Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTCTATATTAAGCTTCGATGAGAGA 140
DB 96 AGCTCTATATTAAGCTTCGATGAGAGA 124

RESULT 8
US-60-202-213-10
; Sequence 10, Application US/60202213
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 38-21(51914)A
; CURRENT APPLICATION NUMBER: US/60/202,213
; CURRENT FILING DATE: 2000-05-08
; NUMBER OF SEQ ID NOS: 17986
; SEQ ID NO 10
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-60-202-213-10

Query Match
Best Local Similarity 100.0%; Score 29; DB 64; Length 385;
Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTCTATATTAAGCTTCGATGAGAGA 140
DB 96 AGCTCTATATTAAGCTTCGATGAGAGA 124
```

```
RESULT 9
US-09-865-439A-18142/C
; Sequence 18142, Application US/09865439A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51936)B
; CURRENT APPLICATION NUMBER: US/09/865,439A
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/207,458
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 18142
; LENGTH: 388
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3206-089-P1-N1-H4
US-09-865-439A-18142

Query Match
Best Local Similarity 100.0%; Score 29; DB 33; Length 388;
Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTCTATATTAAGCTTCGATGAGAGA 140
DB 333 AGCTCTATATTAAGCTTCGATGAGAGA 305

RESULT 10
US-09-565-306-82924
; Sequence 82924, Application US/09565306
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Comer, Timothy W.
; APPLICANT: Lalagudi, Raghunath V.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15459)C
; CURRENT APPLICATION NUMBER: US/09/565,306
; CURRENT FILING DATE: 2000-05-04
; NUMBER OF SEQ ID NOS: 83523
; SEQ ID NO 82924
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: LIB3114-050-P1-K1-G7
US-09-565-306-82924

Query Match
Best Local Similarity 100.0%; Score 29; DB 22; Length 393;
Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTCTATATTAAGCTTCGATGAGAGA 140
DB 110 AGCTCTATATTAAGCTTCGATGAGAGA 138

RESULT 11
US-09-850-147-40
; Sequence 40, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
```



```

; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 40
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-09-850-147-40

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 112 AGCTCTATATTAGCTTCGATGAGAGA 140
DB 135 AGCTCTATATTAGCTTCGATGAGAGA 163

```

```

RESULT 12
US-60-202-213-40
; Sequence 40, Application US/60202213
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51914)A
; CURRENT APPLICATION NUMBER: US/60/202,213
; CURRENT FILING DATE: 2000-05-08
; NUMBER OF SEQ ID NOS: 17986
; SEQ ID NO 40
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-60-202-213-40

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 112 AGCTCTATATTAGCTTCGATGAGAGA 140
DB 135 AGCTCTATATTAGCTTCGATGAGAGA 163

```

```

RESULT 13
US-09-850-147-1092
; Sequence 1092, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1092
; LENGTH: 433
; TYPE: DNA

```

```

; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(433)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-09-850-147-1092

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 112 AGCTCTATATTAGCTTCGATGAGAGA 140
DB 105 AGCTCTATATTAGCTTCGATGAGAGA 133

```

```

RESULT 14
US-60-202-213-1089
; Sequence 1089, Application US/60202213
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51914)A
; CURRENT APPLICATION NUMBER: US/60/202,213
; CURRENT FILING DATE: 2000-05-08
; NUMBER OF SEQ ID NOS: 17986
; SEQ ID NO 1089
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-60-202-213-1089

```

```

Query Match
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 112 AGCTCTATATTAGCTTCGATGAGAGA 140
DB 105 AGCTCTATATTAGCTTCGATGAGAGA 133

```

```

RESULT 15
US-60-207-458-54563/c
; Sequence 54563, Application US/60207458
; GENERAL INFORMATION:
; APPLICANT: Abad, Mark S.
; APPLICANT: Conner, Timothy W.
; APPLICANT: Delkman, Jill
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Laljudi, Raghnath V.
; APPLICANT: Ruan, Yijun G.
; APPLICANT: Ruff, Thomas G.
; APPLICANT: Sammons, R. Douglas
; APPLICANT: Shukla, Hridayabhiranjan
; APPLICANT: Wu, Kunsheng
; APPLICANT: Xu, Nanfei
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51936)A
; CURRENT APPLICATION NUMBER: US/60/207,458
; CURRENT FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 152403
; SEQ ID NO 54563
; LENGTH: 450
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: Clone ID: uc-zmflb731234e12a1
US-60-207-458-54563

```

Query Match 7.2%; Score 29; DB 64; Length 450;
 Best Local Similarity 100.0%; Pred. No. 0.00015;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 112 AGCTTCTATATTAAGCTTCGATGAGAGA 140
 |||||
 Db 321 AGCTTCTATATTAAGCTTCGATGAGAGA 293

Search completed: January 2, 2003, 17:47:03
 Job time : 3761 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 13:24:21 ; Search time 3200 Seconds

(without alignments)
3656.040 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402

Sequence: 1 ccggcgccggcgtaccaccagt.....ggcataacttcctgtgttcc 402

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2054640 seqs, 14551402878 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_com:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_com:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hum:*
31: em_hum:*
32: em_hum:*
33: em_hum:*
34: em_hum:*
35: em_hum:*
36: em_hum:*
37: em_hum:*
38: em_hum:*
39: em_hum:*
40: em_hum:*
41: em_hum:other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	5.7	173852	2 AC095235	AC095235 Rattus no
2	23	5.7	176340	2 AC120750	AC120750 Rattus no
3	20	5.0	369	2 H0M0C0L14	U76577 Human 180 k
4	20	5.0	2964	1 AF063186	AF063186 Arthropac
5	20	5.0	37000	8 SPAC19G12	297209 S.pombe chr
6	20	5.0	59237	2 AC100210	AC100210 Mus muscu
7	20	5.0	60748	2 AC119163	AC119163 Mus muscu
8	20	5.0	84767	9 AC004149	AC004149 Homo sapi
9	20	5.0	128459	2 AC091230	AC091230 Homo sapi
10	20	5.0	141970	9 AL138761	AL138761 Human DNA
11	20	5.0	143620	9 AP001883	AP001883 Homo sapi
12	20	5.0	145695	2 AC015690	AC015690 Homo sapi
13	20	5.0	145870	2 AC068675	AC068675 Homo sapi
14	20	5.0	151937	9 AP001980	AP001980 Homo sapi
15	20	5.0	156186	2 AC036150	AC036150 Homo sapi
16	20	5.0	165315	2 AC073954	AC073954 Homo sapi
17	20	5.0	177744	2 AC109304	AC109304 Mus muscu
18	20	5.0	198446	2 AC123798	AC123798 Mus muscu
19	20	5.0	208979	2 AC023610	AC023610 Mus muscu
20	20	5.0	221576	2 AL731822	AL731822 Mus muscu
21	20	5.0	230160	2 AC020705	AC020705 Homo sapi
22	20	5.0	241585	2 AL844536	AL844536 Mus muscu
23	19	4.7	324	6 AX246754	AX246754 Sequence
24	19	4.7	615	6 AX389622	AX389622 Sequence
25	19	4.7	3519	8 ATH97264	ATH97264 Arabidops
26	19	4.7	3618	8 AF412407	AF412407 Arabidops
27	19	4.7	3786	8 AY096796	AY096796 Arabidops
28	19	4.7	23273	9 AC093389	AC093389 Homo sapi
29	19	4.7	45692	9 AC108153	AC108153 Homo sapi
30	19	4.7	49338	2 AC101240	AC101240 Mus muscu
31	19	4.7	49393	2 AC101083	AC101083 Mus muscu
32	19	4.7	53707	9 AC116359	AC116359 Homo sapi
33	19	4.7	54049	9 AC110089	AC110089 Homo sapi
34	19	4.7	56750	2 AC106014	AC106014 Homo sapi
35	19	4.7	57671	2 AC100231	AC100231 Mus muscu
36	19	4.7	61524	9 AL353725	AL353725 Human DNA
37	19	4.7	69948	2 AC016674	AC016674 Homo sapi
38	19	4.7	75412	2 AC015489	AC015489 Homo sapi
39	19	4.7	81662	8 AB008265	AB008265 Arabidops
40	19	4.7	84952	2 AC015487	AC015487 Homo sapi
41	19	4.7	94414	9 AC051663	AC051663 Homo sapi
42	19	4.7	98339	2 AC114128	AC114128 Rattus no
43	19	4.7	105528	8 AC002392	AC002392 Arabidops
44	19	4.7	109321	9 AC021079	AC021079 Homo sapi
45	19	4.7	114491	9 AC010331	AC010331 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AC095235
DEFINITION Rattus norvegicus clone CH230-10F21, *** SEQUENCING IN PROGRESS
AC095235 173852 bp DNA linear HTG 11-JUL-2002
AC095235 *** 79 unordered pieces.
AC095235 AC095235.3 GI:21722866
VERSION HTG: HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 173852)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbata, J., Benton, J., Blinage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowles, S., Brileva, M., Brown, E., Brown, M., Bryan, N.P.,
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douhwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J., H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvan, J., Kovar, C.,
 Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lileu, C., Liu, J., Liu, W., Louissege, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Mameshwar, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, N., Neal, D., Newton, J., Newtson, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,
 Sodergren, E., Sonaite, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Syatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telirod, B., Thomas, N., Thomas, S.,
 Umanil, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Unpublished
 Direct Submission
 2 (bases 1 to 173852)
 Worley, K.C.
 Direct Submission
 Submitted (16-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 173852)
 Worley, K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17942490.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GDMA
 Center clone name: CH230-10F21
 ----- Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye 100% of reads
 Assembly program: Phrap, version 0.990329
 Consensus quality: 92252 bases at least Q40
 Consensus quality: 103231 bases at least Q30
 Consensus quality: 111005 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 79 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1160: contig of 1160 bp in length
 1260: gap of unknown length
 1261: contig of 1098 bp in length
 2359: gap of unknown length
 2459: contig of 1133 bp in length
 3592: gap of unknown length
 3692: contig of 1171 bp in length
 4863: gap of unknown length
 4963: contig of 1129 bp in length
 6092: gap of unknown length
 6192: contig of 1024 bp in length
 7216: gap of unknown length
 7316: contig of 1166 bp in length
 8482: gap of unknown length
 8582: contig of 1072 bp in length
 9654: gap of unknown length
 9754: contig of 1256 bp in length
 11010: gap of unknown length
 11110: contig of 1148 bp in length
 12257: gap of unknown length
 12357: contig of 1030 bp in length
 13387: gap of unknown length
 13487: contig of 1185 bp in length
 14672: gap of unknown length
 14772: contig of 1031 bp in length
 15803: gap of unknown length
 15903: contig of 1344 bp in length
 17247: gap of unknown length
 17347: contig of 1344 bp in length
 18691: gap of unknown length
 18791: contig of 1129 bp in length
 19920: gap of unknown length
 20020: contig of 1327 bp in length
 21347: gap of unknown length
 21447: contig of 1085 bp in length
 22532: gap of unknown length
 22632: contig of 1015 bp in length
 23647: gap of unknown length
 23747: contig of 1315 bp in length
 25062: gap of unknown length
 25162: contig of 1161 bp in length
 26323: gap of unknown length
 26423: contig of 1568 bp in length
 27991: gap of unknown length
 28091: contig of 1042 bp in length
 29133: gap of unknown length
 29233: contig of 1425 bp in length
 30658: gap of unknown length
 30758: contig of 1194 bp in length
 31952: gap of unknown length
 32052: contig of 1113 bp in length
 33165: gap of unknown length
 33265: contig of 1451 bp in length
 34716: gap of unknown length
 34816: contig of 1870 bp in length
 36686: gap of unknown length
 36787: contig of 1239 bp in length
 38025: gap of unknown length
 38125: contig of 1677 bp in length
 38802: gap of unknown length
 39902: contig of 3008 bp in length
 42910: gap of unknown length
 43010: contig of 1201 bp in length
 44211: gap of unknown length
 44311: contig of 1161 bp in length
 45472: gap of unknown length
 45572: contig of 1161 bp in length

45573 47036: contig of 1464 bp in length
 * 47037 47136: gap of unknown length
 * 47137 48244: contig of 1108 bp in length
 * 48245 48344: gap of unknown length
 * 50213 50213: contig of 1869 bp in length
 * 50214 50313: gap of unknown length
 * 50314 51892: contig of 1579 bp in length
 * 51893 51993: gap of unknown length
 * 51993 53712: contig of 1720 bp in length
 * 53713 53812: gap of unknown length
 * 53813 55724: contig of 1912 bp in length
 * 55725 55824: gap of unknown length
 * 55824 58414: contig of 2590 bp in length
 * 58415 58514: gap of unknown length
 * 58515 60085: contig of 1571 bp in length
 * 60086 60185: gap of unknown length
 * 60186 62192: contig of 2007 bp in length
 * 62193 62292: gap of unknown length
 * 62293 64056: contig of 1764 bp in length
 * 64057 64156: gap of unknown length
 * 64157 66995: contig of 2839 bp in length
 * 66996 67095: gap of unknown length
 * 67096 69950: contig of 2855 bp in length
 * 69951 70050: gap of unknown length
 * 70051 71752: contig of 1702 bp in length
 * 71753 71852: gap of unknown length
 * 71853 73480: contig of 1628 bp in length
 * 73481 73580: gap of unknown length
 * 73581 75880: contig of 2300 bp in length
 * 75881 75980: gap of unknown length
 * 75981 77979: contig of 1999 bp in length
 * 77980 78079: gap of unknown length
 * 78080 79749: contig of 1670 bp in length
 * 79750 79849: gap of unknown length
 * 79850 82131: contig of 2282 bp in length
 * 82132 82231: gap of unknown length
 * 82232 84164: contig of 1933 bp in length
 * 84165 84264: gap of unknown length
 * 84265 86301: contig of 2037 bp in length
 * 86302 86401: gap of unknown length
 * 86402 89582: contig of 3181 bp in length

Query Match 5.7%: Score 23; DB 2; Length 173852;
 Best Local Similarity 100.0%: Pred. No. 0.073;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 149 AGATGTCAGGCTTTCACAGAGA 171
 Db 91195 AGATGTCAGGCTTTCACAGAGA 91217

RESULT 2
 AC120750
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-54113, *** SEQUENCING IN PROGRESS
 *** 62 unordered pieces.
 AC120750 176340 bp DNA linear HTG 24-JUL-2002
 AC120750
 VERSION AC120750.3 GI:21908122
 KEYWORDS HTG; HTGS_Phasel.
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 176340)

REFERENCE
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,I.F., Howard,S., Huber,J., Huylk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvach,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lien,C., Liu,J., Liu,M., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okunolu,G., Orangunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoubokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wellington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Meinstock,G., and Gibbs,R.

Direct Submission
 2 (bases 1 to 176340)
 Unpublished
 Worley,K.C.

Submitted (09-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 176340)
 Worley,K.C.

Direct Submission
 Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Jul 19, 2002 this sequence version replaced g1:20564372.

Genome Center
 Center: Baylor College of Medicine
 Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: GXZE

Center clone name: CH230-54113

Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 127817 bases at least Q40
 Consensus quality: 134179 bases at least Q30
 Consensus quality: 139091 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 NOTE: This is a 'working draft' sequence. It currently
 consists of 62 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

```

* 1 1036: contig of 1036 bp in length
* 1037 1136: gap of unknown length
* 1137 2875: contig of 1739 bp in length
* 2876 2975: gap of unknown length
* 2976 4117: contig of 1142 bp in length
* 4118 4218: gap of unknown length
* 4218 5228: contig of 1011 bp in length
* 5228 5328: gap of unknown length
* 5329 6834: contig of 1506 bp in length
* 6835 8520: contig of 1586 bp in length
* 8521 8620: gap of unknown length
* 8621 10074: contig of 1454 bp in length
* 10075 10175: gap of unknown length
* 10175 11457: contig of 1283 bp in length
* 11458 11557: gap of unknown length
* 11558 12754: contig of 1197 bp in length
* 12755 12854: gap of unknown length
* 12855 14356: contig of 1502 bp in length
* 14357 14456: gap of unknown length
* 14457 15696: contig of 1240 bp in length
* 15697 15796: gap of unknown length
* 15797 17159: contig of 1363 bp in length
* 17160 17259: gap of unknown length
* 17260 18560: contig of 1301 bp in length
* 18561 18660: gap of unknown length
* 18661 19781: gap of unknown length
* 19782 22054: contig of 2273 bp in length
* 22055 22154: gap of unknown length
* 22155 23585: contig of 1431 bp in length
* 23586 25187: contig of 1502 bp in length
* 25188 25287: gap of unknown length
* 25288 26708: contig of 1421 bp in length
* 26709 26808: gap of unknown length
* 26809 27945: contig of 1137 bp in length
* 27946 28045: gap of unknown length
* 28046 29741: contig of 1696 bp in length
* 29742 29841: gap of unknown length
* 29842 31336: contig of 1495 bp in length
* 31337 31436: gap of unknown length
* 31437 33911: contig of 2475 bp in length
* 33912 34011: gap of unknown length
* 34012 35240: contig of 1229 bp in length
* 35241 35340: gap of unknown length
* 35341 36536: contig of 1196 bp in length
* 36537 36636: gap of unknown length
* 36637 38104: contig of 1468 bp in length
* 38105 38204: gap of unknown length
* 38205 40397: contig of 2193 bp in length
* 40398 40497: gap of unknown length
* 40499 42465: contig of 1968 bp in length
* 42466 42565: gap of unknown length
* 42566 43699: contig of 1134 bp in length
* 43700 43799: gap of unknown length
* 43800 45643: contig of 1844 bp in length
* 45644 45743: gap of unknown length
* 45744 47002: contig of 1259 bp in length
* 47003 47102: gap of unknown length
* 47103 48295: contig of 2193 bp in length
* 48296 49395: gap of unknown length
* 49396 51367: contig of 1972 bp in length
* 51368 51467: gap of unknown length
* 51468 53740: contig of 2273 bp in length
* 53741 53840: gap of unknown length
* 53841 56388: contig of 2598 bp in length
* 56389 56439: gap of unknown length
* 56439 59398: contig of 2860 bp in length
* 59399 59498: gap of unknown length
* 59499 62008: contig of 2510 bp in length
* 62009 62108: gap of unknown length
* 62109 63704: contig of 1596 bp in length

```

```

* 63705 63804: gap of unknown length
* 63805 65770: contig of 1966 bp in length
* 65771 65870: gap of unknown length
* 65871 69190: contig of 3320 bp in length
* 69191 69290: gap of unknown length
* 69291 71336: contig of 2046 bp in length
* 71337 71436: gap of unknown length
* 71437 75051: contig of 3615 bp in length
* 75052 75152: gap of unknown length
* 75152 77789: contig of 2638 bp in length
* 77790 77889: gap of unknown length
* 77890 79721: contig of 1832 bp in length
* 79722 79821: gap of unknown length
* 79822 83962: contig of 4141 bp in length
* 83963 84062: gap of unknown length
* 84063 87620: contig of 3558 bp in length
* 87621 87720: gap of unknown length
* 87721 90404: contig of 2684 bp in length
* 90405 90504: gap of unknown length
* 90505 94725: contig of 4221 bp in length
* 94726 94825: gap of unknown length
* 94826 98967: contig of 4142 bp in length
* 98968 99067: gap of unknown length
* 99068 102745: contig of 3678 bp in length
* 102746 102845: gap of unknown length
* 102846 106747: contig of 3901 bp in length
* 106747 106846: gap of unknown length
* 106847 110509: contig of 3663 bp in length
* 110510 110609: gap of unknown length
* 110610 114673: contig of 4064 bp in length
* 114674 114773: gap of unknown length
* 114774 119368: contig of 4594 bp in length
* 119368 123339: gap of unknown length
* 123339: contig of 3872 bp in length

```

Query Match 5.7%; Score 23; DB 2; Length 176340;
 Best Local Similarity 100.0%; Pred. No. 0.073;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 149 AGATGTCAGGCTTCACAGAGA 171
 Db 47941 AGATGTCAGGCTTCACAGAGA 47963

RESIDUE3
 HUMCOL14/c

DEFINITION Human 180 kDa bullous pemphigoid antigen 2/type XVII collagen

ACCESSION HUMCOL14
 VERSION 076577
 KEYWORDS U76577.1 GI:1825501

SEGMENT 14 of 41
 SOURCE Homo sapiens.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 369)
 Gatalica,B., Pulkkinen,L., Li,K., Ryyanen,M.,
 McGrath,J.A. and Uitto,J. Cloning of the human type XVII collagen gene (COL17A1), and
 detection of novel mutations in generalized atrophic benign
 epidermolysis bullosa

AUTHORS Am. J. Hum. Genet. 60 (2), 352-365 (1997)

TITLE JOURNAL MEDLINE
 PUBMED 9012408

REFERENCE 2 (bases 1 to 369)
 Gatalica,B., Pulkkinen,L., Li,K., Ryyanen,M.,
 McGrath,J.A. and Uitto,J. Direct Submission

TITLE JOURNAL Submitted (29-OCT-1996) Dermatology, Thomas Jefferson University,
 233 South 10th Street, Philadelphia, PA 19107, USA

FEATURES Location/Qualifiers

source	1. .369	/organism="Homo sapiens"
	/db_xref="taxon:9606"	
intron	<1. .101	/gene="BPAG2/COL17A1"
	/number=13	
exon	102. .263	/gene="BPAG2/COL17A1"
	/number=14	
intron	264. .369	/gene="BPAG2/COL17A1"
	/number=14	
BASE COUNT	82 a 110 c 76 g 101 t	
ORIGIN		
Query Match	5.0%; Score 20; DB 9; Length 369; *	
Best Local Similarity	100.0%; Pred. No. 5.1;	
Matches	20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	40 GACATGTCGACAGAGAGG 59	
Db	70 GACATGTCGACAGAGAGG 51	
RESULT 4		
AF063186/c	2964 bp DNA linear BCT 02-JAN-2001	
LOCUS	AF063186	
DEFINITION	Archbacter paraffineus DNA polymerase I gene, complete cds.	
ACCESSION	AF063186	
VERSION	AF063186.1 GI:1200209	
KEYWORDS		
SOURCE		
ORGANISM	Rhodococcus erythropolis.	
	Rhodococcus erythropolis.	
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
	Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.	
REFERENCE	1 (bases 1 to 2964)	
AUTHORS	Huang,Y.P. and Ito,J.	
TITLE	Direct Submission	
JOURNAL	Submitted (04-MAY-1998) Microbiology and Immunology, University of Arizona, 1501 N. Campbell Ave., Tucson, AZ 85724, USA	
FEATURES		
source	location/Qualifiers	
	1. 2964	
	/organism="Rhodococcus erythropolis"	
	/strain="ATCC19558"	
	/db_xref="ATCC:19558"	
	/db_xref="taxon:1833"	
	112. 2874	
	/codon_start=1	
	/transl_table=1	
	/product="DNA polymerase I"	
	/protein_id="AAG43148.1"	
	/db_xref="GI:12002010"	
	/translation="MPPVIVYDGSYFLFRAPHALPPLTSTGLTNAIRGASIAIKI MRROPHMAVIEPTPEPTFHEHLSPIYKGRPMPELISQIYPLHALIPLALPLH TLPAEADDDIIGTLAKRAEAHQVLIYSTGRDAAQVTEKVTLEDSEKPELDVGDG FERGVENQNIIDYLTLMGASDGIIMGVPGAKTAALKILEYSGIILENVDIRKIG KVQNIKEHVNDGIALDHOIASIVIDLDFEGYDDIKGEPNIETRLRHLYTEPRNO OSLHPNPNNSNTYQAAKSTINAKAVTAPELDEHSAVTSDDOLGATVYTVLNOD MDTFLERLSTKRRAPETSTSLDYRTAQIYGVSVAPDAQATYVPLAHDEGAPED NREYTLQAIKPILEDSEYKKGHLKTDHYLENHGHLACWYDTMIASTVLASVAT RHGDVDVARYLILSTTYEQVARGAKQKTFNQIPETAHVAEBAHVYRLYEVAU SEKLOADELVNLHNIIEVPAVRLTOMEENGILDLAFLDOLGSEFSNTQNLNED MEIGSEFNSPKOVGEVLEPEKLGKGGKTTGGYSTSEVLEKTEHPITQOLILEF RGLKSTSTYDGLCKQANPDTHRVHTSYHQAALTATGRISTDPNLONIPRAETGRQ IRKAFVAREGVLAADYSQILELMAHLSDEDLALFTHGDVYHRTAAEVGLPEL EEVNDROKRAKAVNFGILGWSFGLRIQUGFROSQDIKQYFHRPEIYDMOR TRVALDGGFETLLGRLYTPDDIDARMVYRKAAERAIINAPLOGSADIKKAMIE VDKMLPKDOAKMLLVHDELVEYDEDAIDELAKLAEVMSVLIQISVPLAVEYKGGK NMDEAH"	
BASE COUNT	823 a 698 c 696 g 747 t	
ORIGIN		
Query Match	5.0%; Score 20; DB 1; Length 2964;	

Best Local Similarity	100.0%	Pred. No. 4.9;			
Matches	20;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
Oy	315	CATATGCCATCTGATGCAT	334		
Db	669	CATATGCCATCTGATGCAT	650		
RESULT 5					
SPAC19G12					
LOCUS		37000 bp	DNA	linear	PLN 12-OCT-1999
DEFINITION	SPAC19G12				
ACCESSION	S.pombe chromosome I cosmid c19G12.				
VERSION	Z97209.1				
KEYWORDS	gi:2879763				
	aldose reductase; carboxypeptidase Y; conserved hypothetical;				
	cut20+; cycloosome/apc subunit cut20/apc4; dna mismatch repair				
	protein; fatty acid hydroxylase; histone h2a-beta; hta2;				
	mitochondrial energy transfer protein; pcyl;				
	phosphatidylinositol-4-phosphate; pmsl; polysaccharide deacetylase;				
	prenyl transferase; RNA-binding; RNP-1; splicing factor;				
	trehalase-phosphatase; tricarboxylate transport; tRNA-Leu;				
	uridylglycolate hydrolase.				
SOURCE	Schizosaccharomyces pombe.				
ORGANISM	Schizosaccharomyces pombe				
	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;				
	Schizosaccharomycetales; Schizosaccharomycetaceae;				
	Schizosaccharomycetes.				
REFERENCE	1 (bases 1 to 37000)				
AUTHORS	Oliver,K., Harris,D., Wood,V., Barrell,B.G. and Rajandream,M.A.				
TITLE	Direct Submission				
JOURNAL	Submitted (30-JUN-1997) Schizosaccharomyces pombe chromosome I				
	sequencing project, Sanger Centre, Wellcome Trust Genome Campus,				
	Hinxton, Camridge CB10 1SA E-mail: barrell@sanger.ac.uk				
COMMENT	On Feb 13, 1998 this sequence version replaced gi:2239193.				
	Notes:				
	Details of yeast sequencing at the Sanger Centre are available on				
	the World Wide Web.				
	(URL, http://www.sanger.ac.uk/Projects/S-pombe/)				
	Protein coding regions (CDS) have been predicted with the help of				
	computer analysis using the Genefinder program in Pombase (an ACEDB				
	database) with additional predictions for the branch-acceptor sites				
	supplied by the program Sp5splice. CAUTION: It is possible that for				
	any individual CDS we may have underestimated or overestimated the				
	number of introns/exons or we may not have chosen the correct				
	splice donor/acceptor sites.				
	CDS are numbered using the following system eg SPAC5H10.01c, SP (S.				
	pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c				
	(complementary strand).				
	The more significant matches with motifs in the PROSITE database				
	are also included but some of these may be fortuitous.				
	The length in codons is given for each CDS.				
	IMPORTANT: This sequence MAY NOT be the entire insert of the				
	sequenced clone. It may be shorter because we only sequence				
	overlapping sections once, or longer, because we arrange for a				
	small overlap between neighbouring subclones.				
	Cosmid c19G12 is overlapped at the 3' end by cosmid c23A1.				
FEATURES	location/Qualifiers				
SOURCE	1. 37000				
	/organism="Schizosaccharomyces pombe"				
	/strain="972h-"				
	/db_xref="taxon:4896"				
	/chromosome="I"				
	/map="IR"				
	/clone="cosmid c19G12"				
	/complement(join(1..823,869..926,1018..1140,1183..1402,				
	1451..1688))				
	/gene="cut20+"				
	/note="SPAC19G12.01c"				
	/complement(join(1..823,869..926,1018..1140,1183..1402,				
	1451..1688))				
	/partial				
	/gene="cut20+"				
	/note="SPAC19G12.01c, len:487"				
CDS					
gene					

```

/codon_start=1
/label=cut20+
/product="cyclosome/apc subunit cut20/apc4"
/protein_id="CAB53521.1"
/db_xref="GI:5791550"
/translation="MWSKSKYKPKKHKREWIETNRVERVSGSLKRVILCPSEMLIAT
LTCSHNLICCRNSQIMDVDFHDLLEAFELCNHDCNLIIVGFKNGEKLITDSGHL
VEQRPASRLAQLMTAMQETIVNEKRNDFLEDAVAPPLIGTLPPSKAEKERSK
AIAQFEPKREGENNKVKELSLIDERRIRYINMESYKIGESDLSKALNLGPIS
HSITNDLHVHLICGNTISLKTLYMDLIGSIDVIATMSTRMCHVLEVL
NAMEEPDNVSEASFITKTPALYSKSDTFESLOLELYOFINAGIPSDLKRMIN
BRVGRVLKNNERAVNSTSLITFCQEVIPACERLTVLLSSARGKSIWGHMKKIP
LDKLVEDCLATGLYLNQNVSEFLNCEERKVMYHPIISMLNVAIVERTSPSSIPP
OEIIEHNETVYIIRHSLFRSKLTSTYFMGTKELOLRD"
complement(join(2169..2404,2446..3702,3774..4239,
4342..4767))
/gene="pms1"
/feature="SPAC19G12.02c"
complement(join(2169..2404,2446..3702,3774..4239,
4342..4767))
/gene="pms1"
/feature="SPAC19G12.02c"
complement(join(2169..2404,2446..3702,3774..4239,
4342..4767))
/gene="pms1"
/codon_start=1
/product="DNA mismatch repair protein pms1"
/protein_id="CAB10113.1"
/db_xref="SWISS-PROT:P54280"
/db_xref="GI:2239195"
/translation="MSTVKPIDANTVHKISGQVITDYASAVKELVENSLOGATTIE
IRKNYGINSTIEVNGSGIDAGDYESTGKKFTKTDPELALQTFPGFGEALSS
ICAVGVITSTRONEAPRGVOLNDHESLSKDLTFPQRTSVAVNDLFCETLVR
KILERNYRESKALSLQAVATISTNKFMYHQTKNSGKLQISTNSNMKALIM
NVETKRVSSSLPMDGILEGYSRPHVSGTSASNERQMLFINRLVMLPKIARYOE
VFPYSAOSPEFALNRLITNGTIDINVSPPKRSVLESEDSIEFKNSLONLCEC
GHAKCSRSOSIESYSOIPDSGSDTOELPOSIPATESETSDSSFSYKSPCKRK
LVEATAOAPISVSVEGASIAQVSKELPELROKDRSSPLNEKVTASSEMKEKLA
LFASSTDSMOKTIDSSFPKOPINKPSSNPNNLLNDPSPASTPYAKTININETS
HNAESVSTLSIPRTQTSVANRIPSKPALOKLFCQSRPLDGLNFKSKTNISLG
VOKDILVSDALKEFNKIGVNIHIDSENOEDLNLTVKRADELRMRVVGOEKRGIIV
VGNMILFIIDHSDSEKFNVEHLKSNLINSODLVPRLDLAATEEVLIDHIDLR
RKGFVALIDNORVGNRCTLSTVPSKTVLIPETSDLETISVLSHPOIDPSSFLER
MLASKACRSVWIGRALTISEMNTIVRHLELSKRWNCPHGPRMRLHLRLKDI"
complement(2405..2417)
/gene="pms1"
/feature="pms1"
/feature="ttaaactctgtag, splice branch and acceptor"
complement(2440..2445)
/gene="pms1"
/feature="gtaagt, splice donor sequence"
complement(join(3623..3702,3774..4239,4342..4758))
/gene="pms1"
/feature="Match to PF01119 DNA_mis-repair, DNA mismatch
repair protein score 529..60"
complement(3703..3719)
/gene="pms1"
/feature="ttaaacaattatttag, splice branch and acceptor"
complement(3768..3773)
/gene="pms1"
/feature="gtaagt, splice donor sequence"
complement(4240..4253)
/gene="pms1"
/feature="ttaaactgtagtag, splice branch and acceptor"
complement(4336..4341)
/gene="pms1"
/feature="gtaagt, splice donor sequence"
complement(4468..4488)
/gene="pms1"
/feature="PS00058 DNA mismatch repair proteins mult"
5410..6372
/gene="SPAC19G12.03"
5410..6372
/gene="SPAC19G12.03"
/feature="SPAC19G12.03"
len:320, LOW SIMILARITY:Yersinia
pseudotuberculosis, CAB46604, hypothetical 33.7 kd
protein.. (293 aa), fasta scores: opt: 155, E():0.003,

```

```

(27.5% identity in 265 aa)"
/codon_start=1
/label=SPAC19G12.03
/product="possible polysaccharide deacetylase"
/protein_id="CAB10114.1"
/db_xref="GI:2239196"
/db_xref="SPTREMBL:O13842"
/translation="MYETRDILGNAGKRPVDTPNPENNSKIAVSFVNYEEGERSLY
EDGEPTLTTEAGLMPFNNRPVRSIESCEYSGSGFWRLINLFKKHYPFCMAI
GOAEKNNVVEAGACEGCEGSHSMINIEGVPEPEYEHIKSYOAIKASPSNS
APRSYTRASLNTKRLVCQYKRDGLQPPDSDBYNDYPMVADPLASRGADDDK
GLIIVPTYLEVNDMAYAVAPGCNSDDTYTARDNFVLYEGEGAKRMKTIGLHCR
LTGRPRERGKQKMEHTISKEGVAVAREQIQAOWSAKHYPKA"
5608..6042
/gene="SPAC19G12.03"
/feature="pms1"
/feature="Pfam match to entry PF01522 Polysac_deacet"
6888..7463
/gene="SPAC19G12.04"
6888..7463
/gene="SPAC19G12.04"
/feature="SPAC19G12.04"
len:191, SIMILARITY:Saccharomyces
cerevisiae, DAL3_YEAST, ureidoglycolate hydrolase, (195
aa), fasta scores: opt: 360, E():5e-17, (36.5% identity in
200 aa)"
/codon_start=1
/label=SPAC19G12.04
/product="putative ureidoglycolate hydrolase"
/protein_id="CAB10115.1"
/db_xref="GI:2239197"
/db_xref="SPTREMBL:O13843"
/db_xref="SPTREMBL:O13843"
/translation="WEARKIYAQALTDEAFAPFGSVQOKDDVKVNSANGTAKYKL
VSESIONEKSSSTSRGKVVNFSTHPSVHPANDENHAFQISVLERPFTQFIPM
CRSSDEQAYLAADNVAAPGCMOTQAFYAKGAGQVYSAGVWHAMVYIGKETML
AAFNENVAVEDDCQVQTESPIEVFKISG"
8397..9272
/gene="SPAC19G12.05"
8397..9272
/gene="SPAC19G12.05"
/feature="SPAC19G12.05"
len:291, SIMILARITY:Saccharomyces
cerevisiae, TXRP_YEAST, tricarboxylate transport protein,
(299 aa), fasta scores: opt: 758, E():0, (49.0% identity in
290 aa)"
/codon_start=1
/label=SPAC19G12.05
/product="putative tricarboxylate transport protein"
/protein_id="CAB10116.1"
/db_xref="GI:2239198"
/db_xref="SPTREMBL:O13844"
/translation="MSTVAKTNPKSSNKPFGVKSIITAGVAGAIESTIYAEFAKTR
LQLRNVEGTAKLPPEGLWYRGCSITVAGNSLKAARFAPFSIKSLSDERGHILR
GPRVLNAGIGAVAESVLPFPESIKTAIIDRRKRPRLKGFLOASRIIVHENGIR
GLYGLIATVAKROANGSVRFYAVNSIKOSQSRLPPEKISTYTTFLVGSVAGITV
YCTQPIDRVKRSMSISASKEKNSIHAKYIILVQDGLLRWMSGATPRLLIISGII
VETYEKMEELKRP"
8397..9020
/gene="SPAC19G12.05"
/feature="pms1"
/feature="Pfam match to entry mito_carr PF00153,
Mitochondrial carrier proteins"
8505..8531
/gene="SPAC19G12.05"
/feature="PS00215 Mitochondrial energy transfer proteins
signature"
complement(10121..10516)
/gene="hta2"

```

Query Match 5.0%; Score 20; DB 8; Length 37000;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 339 TTTCGGTTCTTTGACA 358
|||||
DB 15917 TTTCGGTTCTTTGACA 15936

RESULT 6
AC100210/c 59237 bp DNA linear HTG 22-NOV-2001
LOCUS Mus musculus clone RP23-59M9, LOW-PASS SEQUENCE SAMPLING.
DEFINITION AC100210
AC100210
AC100210.1 GI:17047576
VERSION HTG: HTGS_PHASE0.
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 59237)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-59M9
Unpublished
2 (bases 1 to 59237)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campiano,A., Chang,J., Chazaro,B.,
Choelel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearliano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferrite,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heatford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kells,C., Labroque,K.,
Lamarez,R., Landers,T., Lenocky,J., Levine,R., Liu,G.,
Maclean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Straus,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainou,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center project name: L14347
Center clone name: 59_M_9

* NOTE: This record contains 74 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 693: contig of 693 bp in length
* 694 793: gap of 100 bp
* 794 1492: contig of 699 bp in length
* 1493 1592: gap of 100 bp
* 1593 2294: contig of 702 bp in length
* 2295 2394: gap of 100 bp
* 2395 3040: contig of 646 bp in length
* 3041 3140: gap of 100 bp

3141 3850: contig of 710 bp in length
* 3851 3950: gap of 100 bp
* 3951 4664: contig of 714 bp in length
* 4665 4764: gap of 100 bp
* 4765 5477: contig of 713 bp in length
* 5478 5577: gap of 100 bp
* 5578 6291: contig of 714 bp in length
* 6292 6391: gap of 100 bp
* 6392 7107: contig of 716 bp in length
* 7108 7207: gap of 100 bp
* 7208 7906: contig of 699 bp in length
* 7907 8006: gap of 100 bp
* 8007 8704: contig of 698 bp in length
* 8705 8804: gap of 100 bp
* 8805 9477: contig of 673 bp in length
* 9478 9577: gap of 100 bp
* 9578 10241: contig of 664 bp in length
* 10242 10341: gap of 100 bp
* 10342 11047: contig of 706 bp in length
* 11048 11147: gap of 100 bp
* 11148 11838: contig of 692 bp in length
* 11840 11939: gap of 100 bp
* 11940 12638: contig of 699 bp in length
* 12639 12738: gap of 100 bp
* 12739 13446: contig of 708 bp in length
* 13447 13546: gap of 100 bp
* 13547 14237: contig of 691 bp in length
* 14238 14337: gap of 100 bp
* 14338 15022: contig of 685 bp in length
* 15023 15122: gap of 100 bp
* 15123 15813: contig of 691 bp in length
* 15814 15913: gap of 100 bp
* 15914 16617: contig of 704 bp in length
* 16618 16717: gap of 100 bp
* 16718 17394: contig of 677 bp in length
* 17395 17494: gap of 100 bp
* 17495 18188: contig of 694 bp in length
* 18189 18288: gap of 100 bp
* 18289 18966: contig of 678 bp in length
* 18967 19066: gap of 100 bp
* 19067 19749: contig of 683 bp in length
* 19750 19849: gap of 100 bp
* 19850 20566: contig of 717 bp in length
* 20567 20666: gap of 100 bp
* 20667 21375: contig of 709 bp in length
* 21376 21475: gap of 100 bp
* 21476 22181: contig of 706 bp in length
* 22182 22281: gap of 100 bp
* 22282 22984: contig of 703 bp in length
* 22985 23084: gap of 100 bp
* 23085 23755: contig of 671 bp in length
* 23756 23855: gap of 100 bp
* 23856 24562: contig of 707 bp in length
* 24563 24662: gap of 100 bp
* 24663 25385: contig of 723 bp in length
* 25386 25485: gap of 100 bp
* 25486 26192: contig of 707 bp in length
* 26193 26292: gap of 100 bp
* 26293 26995: contig of 703 bp in length
* 26996 27095: gap of 100 bp
* 27096 27799: contig of 704 bp in length
* 27800 27899: gap of 100 bp
* 27900 28616: contig of 717 bp in length
* 28617 28716: gap of 100 bp
* 28717 29419: contig of 703 bp in length
* 29420 29519: gap of 100 bp
* 29520 30212: contig of 693 bp in length
* 30213 30312: gap of 100 bp
* 30313 31000: contig of 688 bp in length
* 31001 31100: gap of 100 bp
* 31101 31835: contig of 735 bp in length
* 31836 31935: gap of 100 bp
* 31936 32635: contig of 700 bp in length

```

* 32636 32735: gap of 100 bp
* 32736 33445: contig of 710 bp in length
* 33446 33545: gap of 100 bp
* 33546 34256: contig of 711 bp in length
* 34257 34356: gap of 100 bp
* 34357 35057: contig of 701 bp in length
* 35058 35157: gap of 100 bp
* 35158 35865: contig of 708 bp in length
* 35866 35965: gap of 100 bp
* 35966 36670: contig of 705 bp in length
* 36671 36770: gap of 100 bp
* 36771 37576: contig of 806 bp in length
* 37577 37676: gap of 100 bp
* 37677 38391: contig of 715 bp in length
* 38392 38491: gap of 100 bp
* 38492 39208: contig of 717 bp in length
* 39209 39308: gap of 100 bp
* 39309 40017: contig of 709 bp in length
* 40018 40117: gap of 100 bp
* 40118 40827: contig of 710 bp in length
* 40828 40927: gap of 100 bp
* 40928 41612: contig of 685 bp in length
* 41613 41712: gap of 100 bp
* 41713 42406: contig of 694 bp in length
* 42407 42506: gap of 100 bp
* 42507 43203: contig of 697 bp in length
* 43204 43303: gap of 100 bp
* 43304 44000: contig of 697 bp in length
* 44001 44100: gap of 100 bp
* 44101 44807: contig of 707 bp in length
* 44808 44907: gap of 100 bp
* 44908 45615: contig of 708 bp in length
* 45616 45715: gap of 100 bp
* 45716 46420: contig of 705 bp in length
* 46421 46520: gap of 100 bp
* 46521 47227: contig of 707 bp in length
* 47228 47327: gap of 100 bp
* 47328 48044: contig of 717 bp in length
* 48045 48144: gap of 100 bp
* 48145 48846: contig of 702 bp in length
* 48847 48946: gap of 100 bp
* 48947 49652: contig of 706 bp in length
* 49653 49752: gap of 100 bp
* 49753 50447: contig of 695 bp in length
* 50448 50547: gap of 100 bp
* 50548 51225: contig of 678 bp in length
* 51226 51325: gap of 100 bp
* 51326 52026: contig of 701 bp in length
* 52027 52126: gap of 100 bp
* 52127 52844: contig of 718 bp in length
* 52845 52944: gap of 100 bp
* 52945 53649: contig of 705 bp in length
* 53650 53749: gap of 100 bp
* 53750 54463: contig of 714 bp in length
* 54464 54563: gap of 100 bp
* 54564 55259: contig of 696 bp in length
* 55260 55359: gap of 100 bp

```

Query Match 5.0%: Score 20; DB 2; Length 59237;
 Best Local Similarity 100.0%; Pred. No. 4.6;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 AGGCTTAAGACATGTGCCA 50
 ||||||||||||||||
 Db 23149 AGGCTTAAGACATGTGCCA 23130

RESULT 7 AC119163 60748 bp DNA linear HTG 25-APR-2002
 LOCUS AC119163/c
 DEFINITION Mus musculus clone RP24-527P22, LOW-PASS SEQUENCE SAMPLING.
 AC119163
 VERSION AC119163.1 GI:20303744

KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS

TITLE JOURNAL COMMENT

HTG: HTGS_PHASED.
 Mus musculus.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 60748)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Mus musculus, clone RP24-527P22
 Unpublished
 2 (bases 1 to 60748)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
 Boukhgalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
 Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collamore, A.,
 Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galaan, J., Gardyna, S.,
 Ginos, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
 Kanat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
 Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
 Minova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, D.,
 Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
 Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L19420
 Center clone name: 527_P_22

 * NOTE: This record contains 75 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.
 * 1
 * 718 817: gap of 100 bp
 * 818 1532: contig of 715 bp in length
 * 1533 1632: gap of 100 bp
 * 1633 2353: contig of 721 bp in length
 * 2354 2453: gap of 100 bp
 * 2454 3155: contig of 702 bp in length
 * 3156 3255: gap of 100 bp
 * 3256 3961: contig of 706 bp in length
 * 3962 4061: gap of 100 bp
 * 4062 4782: contig of 721 bp in length
 * 4783 4882: gap of 100 bp
 * 4883 5589: contig of 707 bp in length
 * 5590 5689: gap of 100 bp
 * 5690 6395: contig of 706 bp in length

```

* 6396 6495: gap of 100 bp
* 6496 7213: contig of 718 bp in length
* 7214 7313: gap of 100 bp
* 7314 8030: contig of 717 bp in length
* 8031 8130: gap of 100 bp
* 8131 8860: contig of 730 bp in length
* 8861 8960: gap of 100 bp
* 8961 9679: contig of 719 bp in length
* 9680 9780 10515: contig of 736 bp in length
* 10516 10615: gap of 100 bp
* 10616 11349: contig of 734 bp in length
* 11350 11449: gap of 100 bp
* 11450 12165: contig of 716 bp in length
* 12166 12265: gap of 100 bp
* 12266 12978: contig of 713 bp in length
* 12979 13078: gap of 100 bp
* 13079 13772: contig of 694 bp in length
* 13773 13872: gap of 100 bp
* 13873 14587: contig of 715 bp in length
* 14588 14687: gap of 100 bp
* 14688 15433: contig of 746 bp in length
* 15434 15533: gap of 100 bp
* 15534 16238: contig of 705 bp in length
* 16239 16338: gap of 100 bp
* 16339 17065: contig of 727 bp in length
* 17066 17165: gap of 100 bp
* 17166 17871: contig of 706 bp in length
* 17872 17971: gap of 100 bp
* 17972 18655: contig of 684 bp in length
* 18656 18755: gap of 100 bp
* 18756 19450: contig of 695 bp in length
* 19451 19550: gap of 100 bp
* 19551 20277: contig of 727 bp in length
* 20278 20377: gap of 100 bp
* 20378 21094: contig of 717 bp in length
* 21095 21194: gap of 100 bp
* 21195 21738: contig of 544 bp in length
* 21739 21838: gap of 100 bp
* 21839 22533: contig of 695 bp in length
* 22534 22633: gap of 100 bp
* 22634 23350: contig of 717 bp in length
* 23351 23450: gap of 100 bp
* 23451 24172: contig of 722 bp in length
* 24173 24272: gap of 100 bp
* 24273 24968: contig of 697 bp in length
* 24970 25069: gap of 100 bp
* 25070 25785: contig of 716 bp in length
* 25786 25885: gap of 100 bp
* 25886 26596: contig of 711 bp in length
* 26597 26696: gap of 100 bp
* 26697 27413: contig of 723 bp in length
* 27420 27519: gap of 100 bp
* 27520 28247: contig of 728 bp in length
* 28248 28347: gap of 100 bp
* 28348 29064: contig of 717 bp in length
* 29065 29164: gap of 100 bp
* 29165 29875: contig of 711 bp in length
* 29876 29975: gap of 100 bp
* 29976 30686: contig of 711 bp in length
* 30687 30786: gap of 100 bp
* 30787 31500: contig of 714 bp in length
* 31501 31600: gap of 100 bp
* 31601 32302: contig of 702 bp in length
* 32303 32402: gap of 100 bp
* 32403 33107: contig of 705 bp in length
* 33108 33207: gap of 100 bp
* 33208 33919: contig of 712 bp in length
* 33920 34019: gap of 100 bp
* 34020 34732: contig of 713 bp in length
* 34733 34832: gap of 100 bp
* 34833 35558: contig of 726 bp in length
* 35559 35658: gap of 100 bp

```

```

* 35659 36382: contig of 724 bp in length
* 36383 36482: gap of 100 bp
* 36483 37195: contig of 713 bp in length
* 37196 37295: gap of 100 bp
* 37296 37992: contig of 697 bp in length
* 37993 38092: gap of 100 bp
* 38093 38796: contig of 704 bp in length
* 38797 38896: gap of 100 bp
* 38897 39620: contig of 724 bp in length
* 39621 39720: gap of 100 bp
* 39721 40428: contig of 708 bp in length
* 40429 40528: gap of 100 bp
* 40529 41251: contig of 723 bp in length
* 41252 41351: gap of 100 bp
* 41352 42071: contig of 720 bp in length
* 42072 42171: gap of 100 bp
* 42172 42908: contig of 737 bp in length
* 42909 43008: gap of 100 bp
* 43009 43740: contig of 732 bp in length
* 43741 43840: gap of 100 bp
* 43841 44553: contig of 713 bp in length
* 44554 44653: gap of 100 bp
* 44654 45363: contig of 710 bp in length
* 45364 45463: gap of 100 bp
* 45464 46172: contig of 709 bp in length
* 46173 46272: gap of 100 bp
* 46273 46996: contig of 724 bp in length
* 46997 47096: gap of 100 bp
* 47097 47794: contig of 698 bp in length
* 47795 47894: gap of 100 bp
* 47895 48591: contig of 697 bp in length
* 48592 48691: gap of 100 bp
* 48692 49393: contig of 702 bp in length
* 49394 49493: gap of 100 bp
* 49494 50209: contig of 716 bp in length
* 50210 50309: gap of 100 bp
* 50310 51007: contig of 698 bp in length
* 51008 51107: gap of 100 bp
* 51108 51792: contig of 685 bp in length
* 51793 51892: gap of 100 bp
* 51893 52609: contig of 717 bp in length
* 52610 52709: gap of 100 bp
* 52710 53434: contig of 725 bp in length
* 53435 53534: gap of 100 bp
* 53535 54240: contig of 706 bp in length
* 54241 54340: gap of 100 bp
* 54341 55065: contig of 725 bp in length
* 55066 55165: gap of 100 bp
* 55166 55872: contig of 712 bp in length
* 55873 55972: gap of 100 bp

Query Match      5.0%; Score 20; DB 2; Length 60748;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      225 TTGACTTCAGGTGCACATT 244
DB      4315 TTGACTTCAGGTGCACATT 4296
|||||
|||||

RESULT 8
AC0004149/c      84767 bp   DNA   linear   PRI 06-JUN-2000
DEFINITION      Homo sapiens chromosome 17, clone hC17.501_O.10, complete sequence.
ACCESSION       AC0004149
VERSION          AC0004149.1   GI:2981261
KEYWORDS         HTG.
SOURCE           Homo sapiens.
ORGANISM         Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE       1 (bases 1 to 84767)
AUTHORS         Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P. and

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Lander, E.
Homo sapiens chromosome 17, clone hC17.501.O.10
Unpublished
2 (bases 1 to 84767)

Birren, B., Fasman, K., McKernan, K., Nusbaum, C., Richardson, P., Lander, E., Allen, N., Baker, J., Baldwin, J., Barna, N., Beckery, R., Boutwell, C., Byrne, S., Cantu, C., Castle, A., Cerny, J., Cooke, P., Daly, M.J., DePatre, E., Devon, K., Dewar, K., Donelan, L., Durette, B., Ettemad, S., Ferreira, P., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gensheimer, S., Giergery, K., Gilmartin, T., Gray, D., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Linton, L., Mackenzie, J., Marquis, N., McEwan, P., McGuirk, A., Meldrum, J., Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Ranganath, S., Riley, R., Roberts, D., Rollins, G., Rossello, R., Roy, A., Shyam, R., Sochoo, S., Stange-Thomann, N., Stillwell, J., Stone, C., Strickland, C., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wagner, A., Wheeler, T., Wu, T., Ye, W.J., Zemtseva, I., Zhao, J., and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (16-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 84767)

Birren, B., Fasman, K., McKernan, K., Nusbaum, C., Richardson, P., Lander, E., Allen, N., Baker, J., Baldwin, J., Barna, N., Beckery, R., Benn, J., Boatn, C., Boutwell, C., Byrne, S., Cantu, C., Castle, A., Cerny, J., Cooke, P., Daly, M.J., DePatre, E., Devon, K., Dewar, K., Donelan, L., Durette, B., Ettemad, S., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gensheimer, S., Giergery, K., Gilmartin, T., Grant, G., Gray, D., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Kann, L., Linton, L., MacDonald, P., Marquis, N., McEwan, P., McGuirk, A., Meldrum, J., Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rollins, G., Rossello, R., Roy, A., Shyam, R., Stange-Thomann, N., Stillwell, J., Stone, C., Strickland, C., Subramanian, A., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, T., Wu, T., Ye, W.J., Zhao, J., and Zody, M.

TITLE
JOURNAL
REFERENCE
AUTHORS
Submitted (23-MAR-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 84767)

Birren, B., Fasman, K., McKernan, K., Nusbaum, C., Richardson, P., Lander, E., Allen, N., Baker, J., Baldwin, J., Barna, N., Beckery, R., Benn, J., Boatn, C., Boutwell, C., Byrne, S., Cantu, C., Castle, A., Cerny, J., Cooke, P., Daly, M.J., DePatre, E., Devon, K., Dewar, K., Donelan, L., Durette, B., Ettemad, S., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Gensheimer, S., Giergery, K., Gilmartin, T., Grant, G., Gray, D., Hagos, B., Harris, K., Horton, L., Howland, J.C., Hui, L., Jacotot, L., Kann, L., Linton, L., MacDonald, P., Marquis, N., McEwan, P., McGuirk, A., Meldrum, J., Molla, M., Morris, W., Morrow, J., Nachman, A., Naylor, J., O'Connor, T., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Rollins, G., Rossello, R., Roy, A., Shyam, R., Stange-Thomann, N., Stillwell, J., Stone, C., Strickland, C., Subramanian, A., Sydney, K., Tang, L., Vassiliev, H., Vo, A., Wagner, A., Wang, B., Wheeler, T., Wu, T., Ye, W.J., Zhao, J., and Zody, M.

TITLE
JOURNAL
COMMENT
Submitted (06-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 23, 1998 this sequence version replaced g12980974.
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES
SOURCE
Location/Qualifiers
1..84767

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hC17.501.O.10"
/clone_lib="CITC Human BAC"
150..222
/rpt_family="LINE2"
repeat_region

repeat_region complement(289..572)
/rpt_family="AluSx"
repeat_region complement(574..707)
/rpt_family="AluSg/x"
repeat_region complement(2790..2966)
/rpt_family="AluDb"
repeat_region complement(3027..3322)
/rpt_family="AluY"
repeat_region complement(3400..3509)
/rpt_family="MIR"
repeat_region 3623..3699
/rpt_family="(GGCA)n"
repeat_region 4851..4968
/rpt_family="MIR"
repeat_region 5377..5553
/rpt_family="MIR"
repeat_region complement(5725..5910)
/rpt_family="MIR"
repeat_region complement(6877..6906)
/rpt_family="(CA)n"
repeat_region complement(7333..7379)
/rpt_family="MIR"
repeat_region 7386..7668
/rpt_family="AluY"
repeat_region complement(8737..9039)
/rpt_family="AluDb"
repeat_region 9141..9206
/rpt_family="(CA)n"
repeat_region complement(9750..10059)
/rpt_family="AluDb"
repeat_region 10085..10399
/rpt_family="MER58A"
repeat_region complement(10701..10983)
/rpt_family="AluSp"
repeat_region complement(11001..11172)
/rpt_family="AluSg"
repeat_region complement(11173..11473)
/rpt_family="AluSg"
repeat_region complement(11476..11553)
/rpt_family="AluSg"
repeat_region complement(11699..11746)
/rpt_family="MER5B"
repeat_region complement(12281..12397)
/rpt_family="L1MC4"
repeat_region 12415..12715
/rpt_family="AluSx"
repeat_region complement(12761..12931)
/rpt_family="L1MD2"
repeat_region complement(12947..13252)
/rpt_family="AluDb"
repeat_region complement(13253..13622)
/rpt_family="L1ME3"
repeat_region 13788..14096
/rpt_family="AluSg"
repeat_region 14109..14353
/rpt_family="L1"
repeat_region 14354..14851
/rpt_family="MLT2CA"
repeat_region 14852..14990
/rpt_family="L1"
repeat_region 14992..15293
/rpt_family="AluSg"
repeat_region 15307..15899
/rpt_family="L1MB5"
repeat_region complement(15901..16203)
/rpt_family="AluDb"
repeat_region complement(16409..16499)
/rpt_family="L1MC3"
repeat_region complement(16503..16802)
/rpt_family="AluSg"
repeat_region 16959..17266
/rpt_family="AluSx"
repeat_region 18543..18844

```

repeat_region      /rpt_family="AlusX"
18845..19131
complement(19183..19256)
/rpt_family="AluY"
/rpt_family="AluY"
complement(19245..19359)
/rpt_family="AluY"
complement(19509..19679)
/rpt_family="MIR"
20108..20229
/rpt_family="MIR"
20282..20575
/rpt_family="AluO"
complement(20718..20745)
/rpt_family="(GAA)n"
complement(20746..21048)
/rpt_family="AlusX"
complement(21058..21130)
/rpt_family="MIR"
21640..21749
/rpt_family="(CA)n"
complement(22972..23105)
/rpt_family="MIR"
complement(23163..23191)
/rpt_family="(CA)n"
23191..23247
/rpt_family="(GA)n"
23359..23411
/rpt_family="LINE2"
complement(23516..23605)
/rpt_family="MIR"
complement(23623..23732)
/rpt_family="(CA)n"
complement(23800..24071)
/rpt_family="AlusG"
complement(25082..25145)
/rpt_family="(TGGCA)n"
complement(25150..25348)
/rpt_family="LINE2"
complement(25396..25685)
/rpt_family="AlusX"
26709..27007
/rpt_family="AluO"

Query Match      5.0%: Score 20; DB 9; Length 84767;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      339 TTCTGCTTTCTTTGAACA 358
|||||
Db      26525 TTCTGCTTTCTTTGAACA 26506

RESULT 9
AC091230/c      128459 bp      DNA      1linear      HTG 06-AUG-2002
LOCUS      Homo sapiens chromosome 15 clone RP11-414J4 map 15, *** SEQUENCING
IN PROGRESS ***, 4 unordered pieces.
AC091230
AC091230.9      GI:22123234
VERSION      HTG; HTGS_PHASE1; HTGS_FOLDTOP; HTGS_ACTIVEFIN.
KEYWORDS      human.
SOURCE      Homo sapiens
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 128459)
AUTHORS      Birren,B., Nusbaum,C. and Lander,E.
TITLE      Homo sapiens chromosome 15, clone RP11-414J4
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 128459)
AUTHORS      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
Barnett,N., Bastien,V., Boguslavsky,L., Bouhgalter,B., Brown,A.,
Camata,J., Campiano,A., Chang,J., Choepel,Y., Colangelo,M.,

```

```

TITLE      JOURNAL
COMMENT
REFERENCE
AUTHORS
TITLE      JOURNAL
JOURNAL
Submitted (05-APR-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 128459)
Birren,B., Bastien,V., Bloom,T., Boguslavsky,L., Bouhgalter,B.,
Barnett,N., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Camata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Menes,L., Mihova,T., Mienga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Punukhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tefaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:20198593.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WTHR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center Project name: L11927
Center Clone name: 414_J.4
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 68585: contig of 68585 bp in length
* 68586 68685: gap of 100 bp
* 68686 80011: contig of 11326 bp in length
* 80012 80111: gap of 100 bp
* 80112 108353: contig of 28242 bp in length
* 108354 108453: gap of 100 bp
* 108454 128459: contig of 20006 bp in length.
* Location/Qualifiers
1..128459

```

Query Match 5.0%; Score 20; DB 2; Length 128459;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 GATGACAGAGAGAGAGAG 150
|||||
Db 50754 GATGACAGAGAGAGAGAG 50735

RESULT 10
AL138761 141970 bp DNA linear PRI 28-JUL-2000
LOCUS Human DNA sequence from clone Rpl1-16H23 on chromosome 10. Contains
DEFINITION the gene KIAA0204 (HSLK) for a protein kinase, the COL17A1 gene for
collagen type XVII alpha 1 (BP180), ESTs and GSSs, complete
sequence.
ACCESSION AL138761
VERSION AL138761.12 GI:8573811
KEYWORDS HTG; BP180; COL17A1; collagen; HSLK; KIAA0204; protein kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Cniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 141970)
BROWN, A.
Direct Submission
Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequests@sanger.ac.uk
On Jun 20, 2000 this sequence version replaced gi:8452481.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
this a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information
on the WormPep database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormPep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rpl1-16H23 is from the library RPCR-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://daccpac.med.buffalo.edu/
VECTOR: pBAC3.6
This sequence is the entire insert of clone Rpl1-16H23.

FEATURES
Source
1..141970
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rpl1-16H23"
/clone_lib="RCR-11.1"
misc_feature
6..478

repeat_region
41..330
/note="match: GSS: Em: B83625"
/note="LIME1 repeat: matches 5870..6161 of consensus"
331..506
/note="MER20 repeat: matches 9..218 of consensus"
768..1052
/note="AluYo repeat: matches 1..285 of consensus"
2466..2755
/note="AluXx repeat: matches 1..289 of consensus"
3051..3567
/note="LIME repeat: matches 4876..5403 of consensus"
3725..5542
/note="LIMB3 repeat: matches 4371..6182 of consensus"
6046..6250
/note="MER82 repeat: matches 446..653 of consensus"
6285..6488
/note="MIR repeat: matches 2..207 of consensus"
6508..6598
/note="MER82 repeat: matches 2..92 of consensus"
7560..7611
/note="MIR repeat: matches 173..224 of consensus"
7764..7857
/note="L2 repeat: matches 2456..2548 of consensus"
7858..8152
/note="AluXx repeat: matches 1..295 of consensus"
8153..8349
/note="L2 repeat: matches 2548..2750 of consensus"
8412..8728
/note="MER44A repeat: matches 1..333 of consensus"
8988..9065
/note="39 copies 2 mer tt 70% conserved"
9528..10072
/note="MER1A repeat: matches 1..527 of consensus"
10108..10414
/note="AluSg repeat: matches 1..308 of consensus"
11314..11808
/note="MER34 repeat: matches 8..543 of consensus"
11819..11942
/note="MER85 repeat: matches 1..135 of consensus"
11956..11985
/note="MER85 repeat: matches 1..30 of consensus"
12000..12162
/note="MIR repeat: matches 23..198 of consensus"
12296..12415
/note="MIR repeat: matches 18..140 of consensus"
12398..12485
/note="L2 repeat: matches 2611..2697 of consensus"
12514..12899
/note="LIMC1 repeat: matches 5186..5561 of consensus"
12900..13206
/note="AluY repeat: matches 1..307 of consensus"
13207..13912
/note="LIMC1 repeat: matches 5561..6305 of consensus"
13914..13979
/note="LIM4 repeat: matches 5186..5253 of consensus"
14114..52672
/gene="DAl6H23.1"
join(<14114..14278,16037..16085,16423..16572,22338..22410,
22558..22752,23306..23387,24883..25011,25611..26966,
29000..29130,29351..29474,31616..31795,34255..34347,
41599..41728,42223..42247,43173..43361,43927..44052,
45059..45172,49013..52672)
/gene="DAl6H23.1"
/product="DAl6H23.1.2 (protein kinase KIAA0204 (HSLK)
(isoform 2))"
/note="match: CDNAS: Em: D88425 Em: AB015718"
/evidence="not_experimental"
join(<14114..14278,16037..16085,16423..16572,22338..22410,
22558..22752,23306..23387,24883..25011,25611..26966,
29000..29130,29351..29474,31616..31795,34255..41728,
42223..42347,43173..43361,43927..44052,45059..45172,
49013..52672)
/gene="DAl6H23.1"

```

/product="bA16H23.1.1 (protein kinase KIAA0204 (HSLK)
(isoform 1))"
/feature="match: CDNAS: Em:AF112855 Em:AF039574 Em:DB8425
Em:DB6959 Em:AB003357 Em:AB002804
match: ESTS: Em:AI153777 Em:AA667275 Em:AI183019
Em:AW402772 Em:AA190321 Em:AA98865 Em:AL120836 Em:C18798
Em:AA880588 Em:AI048714 Em:AA717749 Em:AA266478 Em:H31623
Em:AI108424 Em:C05749 Em:AA491329 Em:DB2765 Em:AA356218
Em:AW364095 Em:AW604227 Em:AW604224 Em:AW364079 Em:M17289
Em:AI424888 Em:AI558709 Em:ID59390 Em:AA403026 Em:AA702324
Em:AW322697 Em:C15395 Em:AA362324 Em:AI074478 Em:AI128460
Em:AI123337 Em:245955 Em:AA147494 Em:AA398426 Em:AA72565
Em:AI138645 Em:AI378364 Em:AI142068 Em:R88093 Em:C14927
Em:AW024338 Em:AI350740 Em:AA621484 Em:R83865 Em:AW573553
Em:AA333083 Em:RI7527 Em:T36207 Em:AI400071 Em:AA724517
Em:R78860 Em:AA493333 Em:AW300413 Em:AI039339 Em:AI758698
Em:FI3095 Em:AI675770 Em:AA347410 Em:AW604225 Em:H44714
Em:AA307061 Em:H25828 Em:AW021959 Em:W24152 Em:H93694
Em:AA505137 Em:N54648 Em:N58456 Em:AI540244 Em:T75451
Em:AA454970 Em:AA479380 Em:AI206000 Em:AA644445
Em:AI134377 Em:C15275 Em:AI377832 Em:AI949218 Em:AI984184
Em:H09850 Em:AA922539 Em:AA367240 Em:AA922415 Em:AI367042
Em:N78544 Em:N91956 Em:N78186 Em:AI807630 Em:AA772218
Em:AA886401 Em:AA131987 Em:AI004269 Em:AA248981 Em:D60901
Em:AL046816 Em:HA04093 Em:AI205999 Em:AI088423 Em:AA217520
Em:RI2258 Em:R35086 Em:AA129911 Em:AI018333 Em:AA36544
Em:AA454971 Em:AI478214 Em:AI233383 Em:AI949538 Em:N21494
Em:T27233 Em:AI131682 Em:N78047 Em:AI421215 Em:T55950
Em:T39277 Em:N89756 Em:AI134564 Em:AI911945 Em:R86253
Em:AI029638 Em:AA928756 Em:AI124895 Em:N52032 Em:AW325591
Em:AI662844 Em:AI051000 Em:AI613499"
/evidence=not experimental
join(<14114, 14278, 16037, 16085, 16423, 16572, 22338, 22410,
22558, 22752, 23306, 23387, 24883, 25011, 25611, 26966,
29000, 29130, 29351, 29474, 31616, 31795, 34253, 34347,
41599, 41728, 42223, 42347, 43173, 43361, 43927, 44052,
45059, 45172, 49013, 49159)
/gene="bA16H23.1"
/note="isoform based on homology to Guinea pig
serine/threonine kinase (Tr:O55092).
match: proteins: Tr:O55092"
/codon_start=1
/evidence=not experimental
/product="bA16H23.1.2 (protein kinase KIAA0204 (HSLK)
(isoform 2))"
/protein_id="CAC00587.1"
/db_xref="GI:9388136"
/translation="AQNKTSVLAAKVVDTKSEEDLYVVEIDILASCDHPNIVL
LDAYVNNMILIEFCAGAVDAVMELERPPLIESIOVVCQTDALAVLHNKI
HRDLKAGNILEFTDGDIKLADFGVSAKNTIORDSIGFPGYMAEYVWCESKDR
PYDKADVMSLIGTILIEAMIEPHEHLMRYLTKASEPPTLADSPRSSNFKR
LKCKLEKNVARKTTSOLLOHPYTVDSNKPPIRLIAEAKAEVTEVEDGKEDEE
TENSLEIPAKRASSDLSIASSEEDKLSQNACTIESSETERNSNSDKNSKLINK
PTIDEPKAEVEDINEHTIADOLEMTEHLHRAVATKNEERKREKLENDPTEDQEV
DINSVSEKNNINMILETINIENHLKSEEDQEQOMFNKLIKSEIKDTIIQYDV
LVSQETGEKANIQAVDSVGLTEKEDQEQOKDQVINTSVYIGTCEADVA
OKYDEDSAEPTQSDKEVEVVGOKLKNKMWGEGAGTKEVPEKIEVMEIEEGN
KROAINSENIMDINBERGTTEGTTESSSTEMEVRVYADPDQKALSGYODAK
VTTOIDKEKEIPEVSTIKPEPVTVVSOPIEQVPLISININSQSGNKELISLST
ETLIPRESEPKENDNSGTSADTSSIDLNLSISFSLKTSKGSISQETROK
TLKTRKFIYDQGVSVTSKIVTDSISKEELRFLRERLRELFQJKEOROK
NSKLOOQREQIFRRFEQEMSKKROYDELENEKQOKTERLEOCHTNRLREAR
IKGEQEKELSPONMLKNRKKEVINEYERKELMKRKEELRKEELRPRKTLLEEFARK
OKOOOELDSIKTIOOKRAELANIEPCLNNCOQLRABEAAWELFEHIDKQHL
LKQOLQDQYPMORHQLRKREKETBQKQNRQRLIELKAKROTERRARLPRIQSEK
TRMAEKRSLRINSTATPDQDRKIKQFAQEKROKNEMADQKHENMDQLODC
EAVNRELHOLNEKCHLIVHEHOKLELDEHSOELKESR"
/feature="match: ESTS: Em:AF112855 Em:AF039574 Em:DB8425
Em:DB6959 Em:AB003357 Em:AB002804
match: ESTS: Em:AI153777 Em:AA667275 Em:AI183019
Em:AW402772 Em:AA190321 Em:AA98865 Em:AL120836 Em:C18798
Em:AA880588 Em:AI048714 Em:AA717749 Em:AA266478 Em:H31623
Em:AI108424 Em:C05749 Em:AA491329 Em:DB2765 Em:AA356218
Em:AW364095 Em:AW604227 Em:AW604224 Em:AW364079 Em:M17289
Em:AI424888 Em:AI558709 Em:ID59390 Em:AA403026 Em:AA702324
Em:AW322697 Em:C15395 Em:AA362324 Em:AI074478 Em:AI128460
Em:AI123337 Em:245955 Em:AA147494 Em:AA398426 Em:AA72565
Em:AI138645 Em:AI378364 Em:AI142068 Em:R88093 Em:C14927
Em:AW024338 Em:AI350740 Em:AA621484 Em:R83865 Em:AW573553
Em:AA333083 Em:RI7527 Em:T36207 Em:AI400071 Em:AA724517
Em:R78860 Em:AA493333 Em:AW300413 Em:AI039339 Em:AI758698
Em:FI3095 Em:AI675770 Em:AA347410 Em:AW604225 Em:H44714
Em:AA307061 Em:H25828 Em:AW021959 Em:W24152 Em:H93694
Em:AA505137 Em:N54648 Em:N58456 Em:AI540244 Em:T75451
Em:AA454970 Em:AA479380 Em:AI206000 Em:AA644445
Em:AI134377 Em:C15275 Em:AI377832 Em:AI949218 Em:AI984184
Em:H09850 Em:AA922539 Em:AA367240 Em:AA922415 Em:AI367042
Em:N78544 Em:N91956 Em:N78186 Em:AI807630 Em:AA772218
Em:AA886401 Em:AA131987 Em:AI004269 Em:AA248981 Em:D60901
Em:AL046816 Em:HA04093 Em:AI205999 Em:AI088423 Em:AA217520
Em:RI2258 Em:R35086 Em:AA129911 Em:AI018333 Em:AA36544
Em:AA454971 Em:AI478214 Em:AI233383 Em:AI949538 Em:N21494
Em:T27233 Em:AI131682 Em:N78047 Em:AI421215 Em:T55950
Em:T39277 Em:N89756 Em:AI134564 Em:AI911945 Em:R86253
Em:AI029638 Em:AA928756 Em:AI124895 Em:N52032 Em:AW325591
Em:AI662844 Em:AI051000 Em:AI613499"
/evidence=not experimental

```

```

/gene="bA16H23.1"
Query Match 5.0%; Score 20; DB 9; Length 141970;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 40 GACATGTCGACGAGAGG 59
|||||
Db 83751 GACATGTCGACGAGAGG 83770

RESULT 11
AP001883/c
LOCUS
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-140023,
complete sequence.
ACCESSION AP001883
VERSION AP001883
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-140023.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
AUTHORS Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Published only in database (2000)
REFERENCE 2 (bases 1 to 143620)
AUTHORS Hattori, M., Ishii, K., Toyoda, A., Taylor, T. D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou, Tsunumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Mar 16, 2001 this sequence version replaced gi:12381916.
COMMENT Location/Qualifiers
FEATURES
source
1..143620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-140023"
BASE COUNT 41715 a 28357 c 29648 g 43900 t
ORIGIN
Query Match 5.0%; Score 20; DB 9; Length 143620;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 292 GTGTTAATTGGAGGTCA 311
|||||
Db 122317 GTGTTAATTGGAGGTCA 122298

RESULT 12
AC015690/c
LOCUS
DEFINITION Homo sapiens chromosome 11 clone RP11-34A20 map 11, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC015690
VERSION AC015690.3 GI:9107982
KEYWORDS HTG; HTGS_PHRASED.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 145695)
AUTHORS Birten, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 11, clone RP11-34A20
JOURNAL Unpublished

```

REFERENCE
AUTHORS
2 (bases 1 to 145695)
Bairren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckert,R., Boguslavsky,L., Boukhalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donellan,L., Doyle,M.,
Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
Galland,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., MacDonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thoman,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE
JOURNAL
Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced g1:6910854.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

COMMENT
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I1347
Center clone name: 34_A_20

* NOTE: This record contains 163 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1
* 761 860: contig of 760 bp in length
* 861 1636: contig of 776 bp in length
* 1637 1736: gap of 100 bp
* 1737 2494: contig of 758 bp in length
* 2495 2594: gap of 100 bp
* 2595 3361: contig of 767 bp in length
* 3362 3461: gap of 100 bp
* 3462 4246: contig of 785 bp in length
* 4247 4346: gap of 100 bp
* 4347 5140: contig of 794 bp in length
* 5141 5240: gap of 100 bp
* 5241 6025: contig of 785 bp in length
* 6026 6125: gap of 100 bp
* 6126 6928: contig of 803 bp in length
* 6929 7028: gap of 100 bp
* 7029 7788: contig of 760 bp in length
* 7789 7888: gap of 100 bp
* 7889 8671: contig of 783 bp in length
* 8672 8771: gap of 100 bp
* 8772 9531: contig of 760 bp in length
* 9532 9631: gap of 100 bp
* 9632 10429: contig of 798 bp in length
* 10430 10529: gap of 100 bp
* 10530 11304: contig of 775 bp in length
* 11305 11404: gap of 100 bp
* 11405 12183: contig of 779 bp in length
* 12184 12283: gap of 100 bp
* 12284 13075: contig of 792 bp in length
* 13076 13175: gap of 100 bp
* 13176 13973: contig of 798 bp in length

* 13974 14073: gap of 100 bp
* 14074 14866: contig of 793 bp in length
* 14867 14966: gap of 100 bp
* 14967 15748: contig of 782 bp in length
* 15749 15848: gap of 100 bp
* 15849 16628: contig of 780 bp in length
* 16629 16728: gap of 100 bp
* 16729 17512: contig of 784 bp in length
* 17513 17612: gap of 100 bp
* 17613 18443: contig of 831 bp in length
* 18444 18543: gap of 100 bp
* 18544 19306: contig of 763 bp in length
* 19307 19406: gap of 100 bp
* 19407 20176: contig of 770 bp in length
* 20177 20276: gap of 100 bp
* 20277 21055: contig of 779 bp in length
* 21056 21155: gap of 100 bp
* 21156 21932: contig of 777 bp in length
* 21933 22032: gap of 100 bp
* 22033 22812: contig of 780 bp in length
* 22813 22912: gap of 100 bp
* 22913 23702: contig of 790 bp in length
* 23703 23802: gap of 100 bp
* 23803 24592: contig of 790 bp in length
* 24593 24692: gap of 100 bp
* 24693 25487: contig of 795 bp in length
* 25488 25587: gap of 100 bp
* 25588 26377: contig of 790 bp in length
* 26378 26477: gap of 100 bp
* 26478 27289: contig of 812 bp in length
* 27290 27389: gap of 100 bp
* 27390 28171: contig of 782 bp in length
* 28172 28271: gap of 100 bp
* 28272 29053: contig of 782 bp in length
* 29054 29153: gap of 100 bp
* 29154 29937: contig of 784 bp in length
* 29938 30037: gap of 100 bp
* 30038 30815: contig of 778 bp in length
* 30816 30915: gap of 100 bp
* 30916 31688: contig of 773 bp in length
* 31689 31788: gap of 100 bp
* 31789 32555: contig of 767 bp in length
* 32556 32655: gap of 100 bp
* 32656 33437: contig of 782 bp in length
* 33438 33537: gap of 100 bp
* 33538 34337: contig of 800 bp in length
* 34338 34437: gap of 100 bp
* 34438 35186: contig of 749 bp in length
* 35187 35286: gap of 100 bp
* 35287 36083: contig of 797 bp in length
* 36084 36183: gap of 100 bp
* 36184 36967: contig of 784 bp in length
* 36968 37067: gap of 100 bp
* 37068 37852: contig of 785 bp in length
* 37853 37952: gap of 100 bp
* 37953 38723: contig of 771 bp in length
* 38724 38823: gap of 100 bp
* 38824 39574: contig of 751 bp in length
* 39575 39674: gap of 100 bp
* 39675 40456: contig of 782 bp in length
* 40457 40556: gap of 100 bp
* 40557 41319: contig of 763 bp in length
* 41320 41419: gap of 100 bp
* 41420 42175: contig of 756 bp in length
* 42176 42275: gap of 100 bp
* 42276 43048: contig of 773 bp in length
* 43049 43148: gap of 100 bp
* 43149 43907: contig of 759 bp in length
* 43908 44007: gap of 100 bp
* 44008 44805: contig of 798 bp in length
* 44806 44905: gap of 100 bp
* 44906 45698: contig of 793 bp in length
* 45699 45798: gap of 100 bp


```

* 45799 46576: contig of 778 bp in length
* 46577 46576: gap of 100 bp
* 46577 47448: contig of 772 bp in length
* 47449 47548: gap of 100 bp
* 47549 48346: contig of 798 bp in length
* 48347 48446: gap of 100 bp
* 48447 49229: contig of 783 bp in length
* 49230 49329: gap of 100 bp
* 49330 50099: contig of 770 bp in length
* 50100 50199: gap of 100 bp
* 50200 50986: contig of 787 bp in length
* 50987 51086: gap of 100 bp
* 51087 51879: contig of 793 bp in length
* 51880 51979: gap of 100 bp
* 51980 52758: contig of 779 bp in length
* 52759 52858: gap of 100 bp
* 52859 53636: contig of 778 bp in length
* 53637 53736: gap of 100 bp
* 53737 54600: contig of 864 bp in length
* 54601 54700: gap of 100 bp
* 54701 55483: contig of 783 bp in length
* 55484 55583: gap of 100 bp
* 55584 56347: contig of 764 bp in length
* 56348 56447: gap of 100 bp
* 56448 57239: contig of 792 bp in length
* 57240 57339: gap of 100 bp
* 57340 58123: contig of 784 bp in length
* 58124 58223: gap of 100 bp
* 58224 59003: contig of 780 bp in length
* 59004 59103: gap of 100 bp
* 59104 59868: contig of 765 bp in length
* 59869 59968: gap of 100 bp
* 59969 60761: contig of 793 bp in length
* 60762 60861: gap of 100 bp
* 60862 61658: contig of 797 bp in length
* 61659 61758: gap of 100 bp
* 61759 62531: contig of 773 bp in length
* 62532 62631: gap of 100 bp
* 62632 63427: contig of 796 bp in length
* 63428 63527: gap of 100 bp

```

Query Match 5.0%; Score 20; DB 2; Length 145695;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 292 GTGTTAATGTGGAGGTTCA 311
 |||
 Db 107053 GTGTTAATGTGGAGGTTCA 107034

RESULT 13
 AC068675
 LOCUS
 DEFINITION Homo sapiens chromosome 17 clone RP11-39201 map 17, *** SEQUENCING
 IN PROGRESS ***, 4 unordered pieces.
 AC068675
 AC068675.10 GI:22123057
 VERSION
 KEYWORDS HTG; HTGS_PHASeI; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 145670)
 Birren,B., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 17, clone RP11-39201
 Unpublished
 2 (bases 1 to 145670)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
 Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Birren,B., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S.,
 Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,
 Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
 Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
 Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
 Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., MacDonald,P., Major,J., Matthews,C.,
 McCarthy,M., Meldrim,J., Menes,L., Milnova,T., Mieng,V.,
 Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
 O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
 Phunkhand,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Roov,P.,
 Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
 Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
 Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
 Zembek,L., Zimmer,A. and Zody,M.

TITLE
 JOURNAL
 COMMENT
 Submitted (13-AUG-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Aug 6, 2002 this sequence version replaced gi:18543015.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 392_O_1
 Center clone name: LI0324
 NOTE: This is a 'working draft' sequence. It currently
 consists of 4 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 51255: contig of 51255 bp in length
 51256 51355: gap of 100 bp
 51356 55066: contig of 3711 bp in length
 55067 55166: gap of 100 bp
 55167 70769: contig of 15603 bp in length
 70770 70869: gap of 100 bp
 70870 145870: contig of 75001 bp in length.
 Location/Qualifiers
 1. 145870
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="17"

/map="17"
/clone="RP11-39201"
/clone.lib="RP11-11 Human Male BAC"
BASE COUNT 37936 a 35909 c 35452 g 36198 t 375 others
ORIGIN

Query Match 5.0%; Score 20; DB 2; Length 145870;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 339 TTTTCGCTTCTTGACCA 358
Db 21888 TTTTCGCTTCTTGACCA 21907

RESULT 14
AP001980/c
LOCUS AP001980 151937 bp DNA linear PRI 06-MAR-2001
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-264F13,
complete sequence.
ACCESSION AP001980
VERSION AP001980.5 GI:13359356
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-264F13.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
1 (sites)
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (2000)
REFERENCE 2 (bases 1 to 151937)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission

COMMENT
FEATURES
source
Submitted (02-MAR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gs.c.riken.go.jp, URL:http://hgp.gs.c.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Mar 16, 2001 this sequence version replaced gi:12381923.

BASE COUNT 49369 a 28003 c 27962 g 46603 t
ORIGIN
1. 151937
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-264F13"

Query Match 5.0%; Score 20; DB 9; Length 151937;
Best Local Similarity 100.0%; Pred. No. 4.5;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 292 GTGTAAATGTGAGCTTCA 311
Db 10098 GTGTAAATGTGAGCTTCA 10079

RESULT 15
AC036150 156186 bp DNA linear HTG 24-AUG-2002
LOCUS AC036150
DEFINITION Homo sapiens chromosome 11 clone RP11-264F13 map 11, WORKING DRAFT
SEQUENCE, 19 unordered pieces.
ACCESSION AC036150
VERSION AC036150.2 GI:7705190
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP11-264F13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156186)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McKernan,P., McGurk,A., McKernan,K., McPheters,R.,
Meltrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testa,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McKernan,P., McGurk,A., McKernan,K., McPheters,R.,
Meltrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testa,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

COMMENT
TITLE
JOURNAL
Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 156186)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Harford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McKernan,P., McGurk,A., McKernan,K., McPheters,R.,
Meltrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testa,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center project name: 19050
Center clone name: 264_F13
Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 147950 bases at least Q40
Consensus quality: 151905 bases at least Q30
Consensus quality: 153513 bases at least Q20

Insert size: 163000; agarose-fp
 Insert size: 154386; sum-of-contigs
 Quality coverage: 4.4 in Q20 bases; agarose-fp
 Quality coverage: 4.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 19 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 547: contig of 547 bp in length
 * 548 647: gap of 100 bp
 * 648 3155: contig of 2508 bp in length
 * 3156 3255: gap of 100 bp
 * 3256 6384: contig of 3129 bp in length
 * 6385 6484: gap of 100 bp
 * 6485 9217: contig of 2733 bp in length
 * 9218 9317: gap of 100 bp
 * 9318 13050: contig of 3733 bp in length
 * 13051 13150: gap of 100 bp
 * 13151 16192: contig of 3042 bp in length
 * 16193 16292: gap of 100 bp
 * 16293 20043: contig of 3751 bp in length
 * 20044 20143: gap of 100 bp
 * 20144 24398: contig of 4256 bp in length
 * 24400 24499: gap of 100 bp
 * 24500 30398: contig of 5899 bp in length
 * 30399 30498: gap of 100 bp
 * 30499 37678: contig of 7180 bp in length
 * 37679 37778: gap of 100 bp
 * 37779 45560: contig of 7782 bp in length
 * 45561 45660: gap of 100 bp
 * 45661 53184: contig of 7524 bp in length
 * 53185 53284: gap of 100 bp
 * 53285 63459: contig of 10175 bp in length
 * 63460 63559: gap of 100 bp
 * 63560 72171: contig of 8612 bp in length
 * 72172 72271: gap of 100 bp
 * 72272 83976: contig of 11705 bp in length
 * 83977 84076: gap of 100 bp
 * 84077 98526: contig of 14450 bp in length
 * 98527 98626: gap of 100 bp
 * 98627 115880: contig of 17254 bp in length
 * 115881 115980: gap of 100 bp
 * 115981 134149: contig of 18169 bp in length
 * 134150 134249: gap of 100 bp
 * 134250 156186: contig of 21937 bp in length.

FEATURES

source
 1..156186
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="11"
 /map="11"
 /clone="RP11-264F13"
 /clone_lib="RPCT-11 Human Male BAC"
 1..347
 /note="assembly_fragment
 clone_end:SP6
 vector_side:left"
 648..3155
 /note="assembly_fragment"
 3256..6384
 /note="assembly_fragment"
 6485..9217
 /note="assembly_fragment"
 9318..13050
 /note="assembly_fragment"
 13151..16192
 /note="assembly_fragment"
 16293..20043
 /note="assembly_fragment"
 20044..24398
 /note="assembly_fragment"
 24400..24499
 /note="assembly_fragment"
 24500..30398
 /note="assembly_fragment"
 30399..30498
 /note="assembly_fragment"
 30499..37678
 /note="assembly_fragment"
 37679..37778
 /note="assembly_fragment"
 37779..45560
 /note="assembly_fragment"
 45561..45660
 /note="assembly_fragment"
 45661..53184
 /note="assembly_fragment"
 53185..53284
 /note="assembly_fragment"
 53285..63459
 /note="assembly_fragment"
 63460..63559
 /note="assembly_fragment"
 63560..72171
 /note="assembly_fragment"
 72172..72271
 /note="assembly_fragment"
 72272..83976
 /note="assembly_fragment"
 83977..84076
 /note="assembly_fragment"
 84077..98526
 /note="assembly_fragment"
 98527..98626
 /note="assembly_fragment"
 98627..115880
 /note="assembly_fragment"
 115881..115980
 /note="assembly_fragment"
 115981..134149
 /note="assembly_fragment"
 134150..134249
 /note="assembly_fragment"
 134250..156186
 /note="assembly_fragment"

misc_feature /note="assembly_fragment"
 20144..24399
 /note="assembly_fragment"
 misc_feature 24500..30398
 /note="assembly_fragment"
 misc_feature 30499..37678
 /note="assembly_fragment"
 misc_feature 37779..45560
 /note="assembly_fragment"
 misc_feature 45661..53184
 /note="assembly_fragment"
 misc_feature 53285..63459
 /note="assembly_fragment"
 misc_feature 63560..72171
 /note="assembly_fragment"
 misc_feature 72272..83976
 /note="assembly_fragment"
 misc_feature 84077..98526
 /note="assembly_fragment"
 misc_feature 98627..115880
 /note="assembly_fragment"
 misc_feature 115981..134149
 /note="assembly_fragment"
 misc_feature 134250..156186
 /note="assembly_fragment"

BASE COUNT 46555 a 30845 c 30411 g 46571 t 1804 others

ORIGIN

Query Match 5.0%; Score 20; DB 2; Length 156186;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 292 GTGTTAATTGTGAGGTTCA 311
 |||
 Db 105256 GTGTTAATTGTGAGGTTCA 105275

Search completed: January 2, 2003, 16:08:34
 Job time : 3743 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 15:01:56 : Search time 357 Seconds
(Without alignments)
2535.864 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402
Sequence: 1 ccggcgccggcgtaccacgct.....ggcatacttcctgtgttc 402

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 112599159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SID2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SID2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19	4.7	324	22	AAS38626
2	19	4.7	615	24	ABN64583
3	19	4.7	1444	21	AAC46559
4	19	4.7	2016	23	AAS69866
5	19	4.7	2020	23	AAS82153
6	18	4.5	360	20	AAV89959
7	18	4.5	872	21	AAC99130
8	18	4.5	899	22	AAH06889
9	18	4.5	1149	21	AAA51824

C	10	18	4.5	1205	22	AAH14264	Human cDNA sequenc
C	11	18	4.5	1984	21	AAC69571	Human secreted pro
C	12	18	4.5	3099	23	AAS73815	DNA encoding novel
C	13	18	4.5	3099	23	AAS84712	DNA encoding novel
C	14	18	4.5	3137	23	AAS72090	DNA encoding novel
C	15	18	4.5	3288	22	AAH14592	Human cDNA sequenc
C	16	18	4.5	4623	24	ABK09751	Human ovarian tumo
C	17	18	4.5	4633	23	AAS69913	DNA encoding novel
C	18	18	4.5	4689	23	AAS73822	DNA encoding novel
C	19	18	4.5	5136	23	AAS64715	DNA encoding novel
C	20	18	4.5	25277	22	AAS30021	Human lung antigen
C	21	18	4.5	25279	22	AAS28714	Genomic sequence #
C	22	18	4.5	25279	22	AAS30020	Human lung antigen
C	23	18	4.5	25279	22	AAS28713	Genomic sequence #
C	24	18	4.5	30610	22	ABA15643	Human nervous syst
C	25	18	4.5	319608	21	AAH51601	Human chromosome 1
C	26	18	4.5	319608	22	AA509301	Human schizophrant
C	27	17	4.2	179	24	ABN66499	Gene #2997 used to
C	28	17	4.2	179	24	ABN66186	Lung cancer relate
C	29	17	4.2	203	14	AAO60752	Human brain Expres
C	30	17	4.2	254	21	AAC16115	Human secreted pro
C	31	17	4.2	261	21	AAC10644	Human secreted pro
C	32	17	4.2	320	14	AAO60849	Human brain Expres
C	33	17	4.2	324	20	AAK40512	Human secreted pro
C	34	17	4.2	324	21	AAC29123	Human secreted pro
C	35	17	4.2	327	20	AAH86650	Human single nucle
C	36	17	4.2	327	20	AAH87393	Human single nucle
C	37	17	4.2	370	24	ABN20928	Human OREF polynuc
C	38	17	4.2	378	20	AAV87181	EST clone BD489.
C	39	17	4.2	462	22	AAH08928	S. epidermidis ope
C	40	17	4.2	471	22	AAH52447	S. epidermidis ope
C	41	17	4.2	471	22	AAH53375	Human nervous syst
C	42	17	4.2	488	22	ABA13256	Staphylococcus epl
C	43	17	4.2	489	21	ABN06003	Human pancreatic c
C	44	17	4.2	498	21	AAC99118	Human prostate exp
C	45	17	4.2	512	23	ABV50321	

ALIGNMENTS

RESULT 1	
AAS38626	standard; cDNA; 324 BP.
XX	
AC	AAS38626;
XX	
DT	17-DEC-2001 (first entry)
XX	
DE	Novel human diagnostic and therapeutic gene #1684.
XX	
KW	Human; cancer; breast; lung; colon; prostate; cytosolic; diagnostic; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO20016753-A2.
XX	
PD	13-SEP-2001.
XX	
PF	09-MAR-2001; 2001WO-US07787.
XX	
PR	09-MAR-2000; 2000US-0188609.
XX	
PA	(CHIR) CHIRON CORP.
XX	
PA	(HYSE-) HYSED INC.
XX	
PI	Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
XX	Reinhard C, Randazzo F, Kennedy GC, Pot D, Kassam A, Lamson G;
PI	Dzmanac R, Crkventjakov R, Dickson M, Dzmanac S, Labat I;
XX	Leschowitz D, Kita D, Garcia V, Jones WL, Stache-Crain B;
XX	WPI; 2001-530177/58.

PT New polynucleotides and polypeptides, useful for diagnosis and
PT treatment of breast, lung and colon cancer -
XX
PS Claim 1; Page 1018; 1193pp; English.
XX
CC The invention relates to new polynucleotides and polypeptides, useful for
CC diagnosis and treatment of breast, lung and colon cancer. The sequences
CC can be used in detecting differentially expressed genes correlated with a
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample derived from a
CC cell suspected of being cancerous. They can also be used to inhibit
CC tumour growth by modulating expression of a gene product. AAS36943-
CC AAS39338 represent novel human diagnostic and therapeutic coding
CC sequences of the invention.
XX
SQ Sequence 324 BP; 86 A; 59 C; 74 G; 105 T; 0 other;
XX
Query Match 4.7%; Score 19; DB 22; Length 324;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 346 TTCTTTTGACACAGTA 364
DB 205 TTCTTTTGACACAGTA 223
XX
RESULT 2
ABN64583
ID ABN64583 standard; cDNA; 615 BP.
XX
AC ABN64583;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human cancer related polynucleotide SEQ ID NO 4550.
XX
KW Human; cytostatic; gene expression; gene mapping; tissue profiling;
KW gene therapy; cancer; tumour; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200214500-A2.
XX
PD 21-FEB-2002.
XX
PF 16-AUG-2001; 2001WO-US25840.
XX
PR 16-AUG-2000; 2000US-226326P.
XX
PA (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;
PI Lamson G, Scott EM, Zhang G, Kassam A, Fot D, Labat I;
XX
DR WPI: 2002-241905/29.
XX
PT New nucleic acid for producing a polypeptide, detecting differentially
PT expressed genes correlated with a cancerous state of a mammalian cell,
PT and inhibiting tumor growth -
XX
PS Claim 1; SEQ ID NO 4550; 883pp + Sequence Listing; English.
XX
CC The invention relates to an isolated polynucleotide (ABN27253-ABN33262)
CC with cytostatic activity. The polynucleotide is used to produce a
CC polypeptide, to detect differentially expressed genes correlated with a
CC cancerous state of a mammalian cell and to inhibit tumour growth. The
CC polynucleotide is used as a probe in mapping and tissue profiling. The
CC encoded polypeptide and antibodies to the polypeptide can also be used
CC for therapeutic and diagnostic purposes. The polynucleotide is useful for
CC gene therapy.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 615 BP; 181 A; 123 C; 149 G; 162 T; 0 other;
XX
Query Match 4.7%; Score 19; DB 24; Length 615;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 346 TTCTTTTGACACAGTA 364
DB 205 TTCTTTTGACACAGTA 223
XX
RESULT 3
AAC46559
ID AAC46559 standard; DNA; 1444 BP.
XX
AC AAC46559;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 50583.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 16-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159332.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 4.7%; Score 19; DB 21; Length 1444;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 122 TTAACTTCATGAGAAGA 140
|||||
DB 1121 TTAACTTCATGAGAAGA 1139

RESULT 4
AAS69866/c
ID AAS69866 standard; cDNA; 2016 BP.
XX

AC AAS69866;
XX

DT 13-FEB-2002 (first entry)
XX

DE DNA encoding novel human diagnostic protein #5670.
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX

OS Homo sapiens.
XX

PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
XX

PR 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI: 2001-639362/73.
XX

DR P-PSDB; ABG05679.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1: SEQ ID NO 5670; 103bp; English.
XX

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantifying a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 2016 BP; 338 A; 631 C; 608 G; 439 T; 0 other;

Query Match 4.7%; Score 19; DB 23; Length 2016;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 AGACATGTCCAGAGAAG 57

DB 2004 AGACATGTCCAGAGAAG 1986
|||||

RESULT 5
AAS82153
ID AAS82153 standard; cDNA; 2020 BP.
XX

AC AAS82153;
XX

DT 13-FEB-2002 (first entry)
XX

DE DNA encoding novel human diagnostic protein #17957.
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX

OS Homo sapiens.
XX

PN WO200175067-A2.
XX

PD 11-OCT-2001.
XX

PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
XX

PR 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI: 2001-639362/73.
XX

DR P-PSDB; ABG17966.
XX

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 1: SEQ ID NO 17957; 103bp; English.
XX

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantifying a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 2020 BP; 439 A; 610 C; 634 G; 337 T; 0 other;

Query Match 4.7%; Score 19; DB 23; Length 2020;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 39 AGACATGTCCAGAGAAG 57
|||||

DB 13 AGACATGTCCAGAGAAG 31

RESULT 6
AAV89959
ID AAV89959 standard; cDNA; 360 BP.
XX
AC AAV89959;
XX
DT 15-FEB-1999 (first entry)
XX
DE EST clone CS765.
XX
KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW tissue growth; actinin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO9845436-A2.
XX
PD 15-OCT-1998.
XX
PF 10-APR-1998; 98WO-US06955.
XX
PR 10-APR-1997; 97US-0838821.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX
DR WPI; 1999-070077/06.
XX
PT New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries.
XX
PS Claim 1; Page 386; 618pp; English.
XX
CC The present sequence represents a human expressed sequence tag (EST).
CC The polynucleotide, which is a secreted EST, and the encoded protein
CC are predicted to have useful biological activities which would make
CC them suitable for treating, preventing or ameliorating medical
CC conditions in humans and animals, although no supporting data is
CC given. Suggested activities include nutritional activity, immune
CC stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, actinin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC activity, receptor/ligand activity, anti-inflammatory activity,
CC cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The polynucleotide may also be useful for gene therapy.
XX
SQ Sequence 360 BP; 124 A; 79 C; 95 G; 62 T; 0 other;
XX
Query Match 4.5%; Score 18; DB 20; Length 360;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 134 GAGAGAAGAGAAGAGA 151
|||
Db 283 GAGAGAAGAGAAGAGA 300

RESULT 7
AAC99130
ID AAC99130 standard; cDNA; 872 BP.
XX
AC AAC99130;
XX
DT 09-MAR-2001 (first entry)
XX
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:358.

XX
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW detection; diagnosis; identification; cytostatic; neuroprotective;
KW neurotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW antiinflammatory; cardiac; gene therapy; chromosome mapping;
KW linkage analysis; tissue identification; tissue typing; forensic;
KW neural; immune system; muscular; reproductive; gastrointestinal;
KW pulmonary; cardiovascular; renal; proliferative; ss.
XX
OS Homo sapiens.
XX
PN WO200055320-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05989.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI; 2000-579444/54.
DR P-PSDB; AAB54365.
XX
PT New nucleic acid that is a pancreatic cancer antigen for preventing,
PT treating, or ameliorating a medical condition, particular pancreatic
PT cancer, or for use in assays for diagnosing a pathological condition -
XX
PS Claim 1; Page 780; 1379pp; English.
XX
CC AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC proteins, called pancreatic cancer antigens, given in AAB54008 to
CC AAB54466. The human pancreatic cancer antigens have cytostatic,
CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,
CC gynaecological, cardiac and antiinflammatory activities, and can be used
CC in gene therapy. The polynucleotide and proteins can be used for
CC preventing, treating, or ameliorating a medical condition or in assays
CC for diagnosing a pathological condition or a susceptibility to one in a
CC subject. Binding partners to the proteins and the activity of the
CC proteins can be identified. The pancreatic cancer antigens can be used to
CC detect, treat or prevent pancreatic disorders, especially cancer.
CC Agonists and antagonists to the antigens can be screened for. The
CC pancreatic cancer antigen polynucleotides can be used to design nucleic
CC acid hybridisation probes that can be used in chromosome mapping, linkage
CC analysis, tissue identification and/or typing and a variety of forensic
CC and diagnostic methods. The proteins can be used to generate antibodies
CC which are used to purify, detect and target the polypeptides, including
CC both in vivo and in vitro diagnostic and therapeutic methods. The
CC proteins can be used to treat or prevent neural, immune system, muscular,
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 872 BP; 253 A; 186 C; 244 G; 183 T; 6 other;
XX
Query Match 4.5%; Score 18; DB 21; Length 872;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 134 GAGAGAAGAGAAGAGA 151
|||
Db 468 GAGAGAAGAGAAGAGA 485

RESULT 8
AAH06889
ID AAH06889 standard; cDNA; 899 BP.
XX
AC AAH06889;
XX
DT 26-JUN-2001 (first entry)

```

XX DE Human cDNA clone (5'-primer) SEQ ID NO:3724.
XX XX
XX KM Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX OS Homo sapiens.
XX PN EPI074617-A2.
XX PD 07-FEB-2001.
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA (HELI-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WPI; 2001-318749/34.
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs -
XX PS Claim 1; SEQ ID 3724; 2537pp + CD ROM; English.
XX CC The present invention describes primer sets for synthesizing 5602
XX CC full-length cDNAs defined in the specification. Where a primer set
XX CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX CC to the complementary strand of a polynucleotide which comprises one of
XX CC the 5602 nucleotide sequences defined in the specification, where the
XX CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX CC of an oligonucleotide comprising a sequence complementary to the
XX CC complementary strand of a polynucleotide which comprises a 5'-end
XX CC sequence and an oligonucleotide comprising a sequence complementary to a
XX CC polynucleotide which comprises a 3'-end sequence, where the
XX CC oligonucleotide comprises at least 15 nucleotides and the combination of
XX CC the 5'-end sequence/3'-end sequence is selected from those defined in
XX CC the specification. The primer sets can be used in antisense therapy and
XX CC in gene therapy. The primers are useful for synthesizing polynucleotides,
XX CC particularly full-length cDNAs. The primers are also useful for the
XX CC detection and/or diagnosis of the abnormality of the proteins encoded by
XX CC the full-length cDNAs. The primers allow obtaining of the full-length
XX CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX CC represent oligonucleotides, all of which are used in the exemplification
XX CC of the present invention.
XX SQ Sequence 899 BP; 246 A; 205 C; 265 G; 179 T; 4 other;
XX
XX Query Match 4.5%; Score 18; DB 22; Length 899;
XX Best Local Similarity 100.0%; Pred. No. 48;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 134 GAGAGAGAGAGAGAGA 151
DB 742 GAGAGAGAGAGAGAGA 759

```

```

RESULT 9
AAAS1824
ID AAAS1824 standard; cDNA; 1149 BP.
XX AC AAAS1824;
XX KM
XX

```

```

DT 09-JAN-2001 (first entry)
XX DE Human melanocortin 1 receptor protein MC-R1ESTc1 cDNA.
XX KM MC-R1; melanocortin 1 receptor; rhodopsin; G-protein coupled receptor;
XX KM splice variant; MC-R1b; C-terminal extension; agonist; antagonist;
XX KM modulator; ss.
XX OS Homo sapiens.
XX PF Key Location/Qualifiers
XX FT 1..1149
XX FT CDS /tag= a
XX FT /product= Melanocortin_1_receptor_B_splice_variant
XX PN W020039147-A1.
XX PD 06-JUL-2000.
XX PR 16-DEC-1999; 99WO-US29963.
XX PR 23-DEC-1998; 98US-0113401.
XX PA (MERI ) MERCK & CO INC.
XX PI Howard AD, MacNeil DJ, Van Der Ploeg LHT;
XX DR WPI; 2000-452365/39.
XX DR P-PSDB: AAY97019.
XX PT New nucleic acid encoding a human melanocortin 1 receptor protein
XX PT (MC-R1b) for determining whether a substance is capable of binding to
XX PT or activating human MC-R1b and identifying a substance that modulates
XX PT MC-R1b receptor activity
XX PS Claim 2; Page 73; 101pp; English.
XX CC The invention concerns novel splice variants of the melanocortin 1
XX CC receptor (MC-R1) protein belonging to the rhodopsin sub-family of
XX CC G-protein coupled receptors. The splice variants, referred to as MC-R1b
XX CC proteins, contain an intracellular domain with an additional 65 amino
XX CC acid residues in comparison to previously disclosed human MC-R1, referred
XX CC to as MC-R1a. Additionally, residue 317 of the MC-R1b proteins is Cys,
XX CC whereas the C-terminal amino acid residue 317 of known MC-R1a proteins is
XX CC Trp. The novel sequences can be used to determine whether a substance
XX CC modulates MC-R1b receptor activity and to identify potential agonists or
XX CC antagonists of MC-R1b. Pharmaceutical compositions comprising
XX CC modulators of MC-R1b are used to treat or diagnose disorders involving
XX CC inappropriate melanocortin expression or activity.
XX SQ Sequence 1149 BP; 188 A; 402 C; 319 G; 240 T; 0 other;
XX
XX Query Match 4.5%; Score 18; DB 21; Length 1149;
XX Best Local Similarity 100.0%; Pred. No. 48;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 48 CCAGGAGAGGGCCCTTG 65
DB 1032 CCAGGAGAGGGCCCTTG 1049

```

```

RESULT 10
AAH14264/C
ID AAH14264 standard; cDNA; 1205 BP.
XX AC AAH14264;
XX KM
XX DT 26-JUN-2001 (first entry)
XX DE Human cDNA sequence SEQ ID NO:11581.
XX KM Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
XX

```

OS Homo sapiens.
 XX EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 XX 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-ANG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 XX Claim 8; SEQ ID 11581; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any special methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SO Sequence 1205 BP; 313 A; 283 C; 309 G; 300 T; 0 other;
 Query Match 4.5%; Score 18; DB 22; Length 1205;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 48 CCAGGAGAGAGGCGCTTG 65
 DB 599 CCAGGAGAGAGGCGCTTG 582
 RESULT 11
 AAC69571/C
 ID AAC69571 standard; DNA; 1984 BP.
 XX
 AC AAC69571;
 XX
 DT 31-JAN-2001 (first entry)
 XX
 DE Human secreted protein gene 60 clone HMSGK61.
 XX
 KW Immunosuppressive; antiarthritic; antirheumatic; antiproliferative;

KW cytostatic; cardiant; vasotropic; cerebroprotective; neuroprotective;
 KW neotropic; antibacterial; virucide; fungicide; ophthalmological;
 KW vulnerary; gene therapy; infection; human; secreted protein; ss.
 XX
 OS Homo sapiens.
 XX
 XX WO200061623-A1.
 XX
 PD 19-OCT-2000.
 XX
 XX 06-APR-2000; 2000MO-US08979.
 XX
 PR 09-APR-1999; 99US-0128693.
 PR 26-APR-1999; 99US-0130991.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Ruben SM, NI J, Komatsoulis GA, Rosen CA, Soppet DR, Shi Y;
 PI Lafleur DW, Olsen HS, Ebner R, Florence KA, Moore PA, Birse CE;
 PI Young PE;
 XX
 DR WPI; 2000-647418/62.
 XX
 PT New nucleic acid molecules encoding 62 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS
 XX Claim 1; Page 574-575; 716pp; English.
 XX
 CC The invention relates to the isolation of genes (AAC69512-C69587)
 CC encoding 62 human secreted proteins (AAB38321-B38396). The genes can be
 CC used to generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC69503) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) autoimmune diseases e.g.
 CC rheumatoid arthritis; (b) hyperproliferative disorders e.g. neoplasms
 CC of the breast or liver; (c) cardiovascular disorders e.g. cardiac
 CC arrest; (d) cerebrovascular disorders e.g. cerebral ischemia; (e)
 CC angiogenesis; (f) nervous system disorders e.g. Alzheimer's disease; (g)
 CC infections caused by bacteria, viruses and fungi; and (h) ocular
 CC disorders e.g. corneal infection. The polypeptides can also be used to
 CC aid wound healing and epithelial cell proliferation, to prevent skin
 CC aging due to sunburn, to maintain organs before transplantation, for
 CC supporting cell culture of primary tissues, to regenerate tissues and in
 CC chemotaxis.
 XX
 SO Sequence 1984 BP; 422 A; 512 C; 440 G; 610 T; 0 other;
 Query Match 4.5%; Score 18; DB 21; Length 1984;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 134 GAGAGAGAGAGAGAGAGA 151
 DB 1090 GAGAGAGAGAGAGAGAGA 1073
 RESULT 12
 AAS73815
 ID AAS73815 standard; cDNA; 3099 BP.
 XX
 AC AAS73815;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #9619.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

```

XX Homo sapiens.
OS
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB: ABG09628.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 9619; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3099 BP; 881 A; 794 C; 818 G; 606 T; 0 other;
XX
XX Query Match 4.5%; Score 18; DB 23; Length 3099;
XX Best Local Similarity 100.0%; Pred. No. 47;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 134 GAGAGAGAGAGAGAGA 151
XX ||||||||||||||||
XX Db 724 GAGAGAGAGAGAGAGA 741
XX
XX RESULT 13
XX AAS84712
XX ID AAS84712 standard; cDNA; 3099 BP.
XX
XX AAS84712;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #20516.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX

```

```

XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX
XX P-PSDB: ABG20525.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 1; SEQ ID No 20516; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAS64197-AAS94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 3099 BP; 881 A; 794 C; 818 G; 606 T; 0 other;
XX
XX Query Match 4.5%; Score 18; DB 23; Length 3099;
XX Best Local Similarity 100.0%; Pred. No. 47;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 134 GAGAGAGAGAGAGAGA 151
XX ||||||||||||||||
XX Db 724 GAGAGAGAGAGAGAGA 741
XX
XX RESULT 14
XX AAS72090
XX ID AAS72090 standard; cDNA; 3137 BP.
XX
XX AAS72090;
XX
XX 13-FEB-2002 (first entry)
XX
XX DNA encoding novel human diagnostic protein #7894.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX

```

```

XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR
XX P-PSDB; ABC07903.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID NO 7894; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AA64197-AAA94564 represent novel human
XX diagnostic coding sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 3137 BP; 886 A; 791 C; 840 G; 620 T; 0 other;
XX
XX Query Match 4.5%; Score 18; DB 23; Length 3137;
XX Best Local Similarity 100.0%; Pred. No. 47;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 134 GAGAGAAGGAAGGAGA 151
Db 724 GAGAGAAGGAAGGAGA 741

```

```

RESULT 15
AAH14592
ID AAH14592 standard; cDNA; 3288 BP.
XX
XX AAH14592;
AC
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX Human cDNA sequence SEQ ID NO:12199.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
OS
XX
XX BP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX

```

```

PE 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
PR
XX 27-AUG-1999; 99JP-0300253.
PR
XX 11-JAN-2000; 2000JP-0118776.
PR
XX 02-MAY-2000; 2000JP-0183767.
PR
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELT-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
PS Claim 8; SEQ ID 12199; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
SQ Sequence 3288 BP; 945 A; 768 C; 882 G; 693 T; 0 other;
XX
XX Query Match 4.5%; Score 18; DB 22; Length 3288;
XX Best Local Similarity 100.0%; Pred. No. 47;
XX Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 134 GAGAGAAGGAAGGAGA 151
Db 742 GAGAGAAGGAAGGAGA 759

```

```

Search completed: January 2, 2003, 16:07:17
Job time : 363 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 14:25:11 ; Search time 75 Seconds
(without alignments)
1643.787 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccggcgccggcgtaccacagct.....gycataactctgtgtcc 402

Scoring table: OLIGO_MNC
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_MA:*
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	4.2	489	4	US-09-134-001C-66 Sequence 66, Appl
2	17	4.2	597	4	US-09-134-001C-1790 Sequence 1790, Ap
3	17	4.2	1393	4	US-09-737-698B-11 Sequence 11, Appl
4	17	4.2	1766	5	PCT-US92-06532-2 Sequence 2, Appl
5	17	4.2	1946	4	US-09-737-698B-27 Sequence 27, Appl
6	16	4.0	40	2	US-08-857-946-146 Sequence 146, App
7	16	4.0	40	3	US-08-970-740-146 Sequence 146, App
8	16	4.0	99	1	US-07-918-953-3 Sequence 3, Appl
9	16	4.0	99	1	US-08-081-661-3 Sequence 3, Appl
10	16	4.0	242	1	US-08-487-001A-32 Sequence 32, Appl
11	16	4.0	242	2	US-08-630-822A-32 Sequence 32, Appl
12	16	4.0	242	2	US-09-005-069-32 Sequence 32, Appl
13	16	4.0	275	1	US-07-918-953-12 Sequence 12, Appl
14	16	4.0	275	1	US-07-918-953-14 Sequence 14, Appl
15	16	4.0	275	1	US-08-081-661-12 Sequence 12, Appl
16	16	4.0	275	1	US-08-081-661-14 Sequence 14, Appl
17	16	4.0	275	1	US-07-918-953-16 Sequence 16, Appl
18	16	4.0	276	1	US-08-081-661-16 Sequence 16, Appl
19	16	4.0	298	1	US-07-826-928A-28 Sequence 28, Appl
20	16	4.0	320	4	US-08-994-962-14 Sequence 14, Appl
21	16	4.0	330	5	PCT-US95-08596-1 Sequence 1, Appl
22	16	4.0	359	3	US-08-589-028-3 Sequence 3, Appl
23	16	4.0	359	3	US-08-784-582-3 Sequence 3, Appl
24	16	4.0	359	4	US-08-785-271-3 Sequence 3, Appl
25	16	4.0	416	4	US-08-945-140-5 Sequence 5, Appl
26	16	4.0	457	4	US-08-994-962-12 Sequence 12, Appl
27	16	4.0	510	1	US-07-918-953-7 Sequence 7, Appl

c 28	16	4.0	510	1	US-08-081-661-7	Sequence 7, Appl
c 29	16	4.0	514	4	US-08-994-962-11	Sequence 11, Appl
c 30	16	4.0	515	3	US-08-589-028-1	Sequence 1, Appl
c 31	16	4.0	515	3	US-08-784-582-1	Sequence 1, Appl
c 32	16	4.0	515	4	US-08-785-271-1	Sequence 1, Appl
c 33	16	4.0	515	4	US-08-537-696-11	Sequence 11, Appl
c 34	16	4.0	633	4	US-09-537-696-12	Sequence 12, Appl
c 35	16	4.0	666	4	US-09-537-696-13	Sequence 13, Appl
c 36	16	4.0	674	3	US-09-069-811-1	Sequence 1, Appl
c 37	16	4.0	674	4	US-09-543-106-1	Sequence 1, Appl
c 38	16	4.0	860	4	US-08-994-962-1	Sequence 1, Appl
c 39	16	4.0	1131	1	US-07-959-946-2	Sequence 2, Appl
c 40	16	4.0	1131	1	US-08-333-577-2	Sequence 2, Appl
c 41	16	4.0	1131	5	PCT-US92-08634-2	Sequence 2, Appl
c 42	16	4.0	1766	4	US-09-517-467B-18	Sequence 18, Appl
c 43	16	4.0	2492	3	US-08-655-191-3	Sequence 3, Appl
c 44	16	4.0	2492	3	US-08-682-080-3	Sequence 3, Appl
c 45	16	4.0	2659	3	US-09-029-267-19	Sequence 19, Appl

ALIGNMENTS

```

RESULT 1
US-09-134-001C-66
; Sequence 66, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GNC-007
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 66
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-66

Query Match      4.2%; Score 17; DB 4; Length 489;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 223 CTTGACTTCAGGTGGA 239
Db 332 CTTGACTTCAGGTGGA 348

RESULT 2
US-09-134-001C-1790
; Sequence 1790, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; FILE REFERENCE: GNC-007
; EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 1790
; LENGTH: 597
; TYPE: DNA

```

ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1790

Query Match 4.2%; Score 17; DB 4; Length 597;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 AGGCATATTATAATCT 266
|||||
DB 503 AGGCATATTATAATCT 519

RESULT 3
US-09-737-698B-11/c
Sequence 11, Application US/09737698B

Patent No. 6462258
GENERAL INFORMATION:
APPLICANT: Fincher, Karen
APPLICANT: Wilkinson, Jack
TITLE OF INVENTION: No. 6462258el Plant Expression Constructs
FILE REFERENCE: 38-21(51499)C
CURRENT APPLICATION NUMBER: US/09/737,698B
CURRENT FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: US 60/171,173
PRIOR FILING DATE: 1999-12-16
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn version 3.0
SEQ ID NO 11
LENGTH: 1393
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1393)
OTHER INFORMATION: Act11 promoter polynucleotide sequence and intron
US-09-737-698B-11

Query Match 4.2%; Score 17; DB 4; Length 1393;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 TCTGATGATTCATTT 341
|||||
DB 30 TCTGATGATTCATTT 14

RESULT 4
PCT-US92-06532-2
Sequence 2, Application PC/TUS9206532

GENERAL INFORMATION:
APPLICANT: Krause, James F.
TITLE OF INVENTION: Human Substance P Receptor
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scott J. Meyer, Monsanto Co., A3SD
STREET: 800 N. Lindbergh Blvd.
CITY: St. Louis
STATE: Missouri
COUNTRY: U.S.A
ZIP: 63167
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06532
FILING DATE: 19920805
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meyer, Scott J.
REGISTRATION NUMBER: 25,275
REFERENCE/DOCKET NUMBER: 07-24(776)A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (314)694-3117
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 1766 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 211..1431
PCT-US92-06532-2

Query Match 4.2%; Score 17; DB 5; Length 1766;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 329 ATGCATTCATTTCTGG 345
|||||
DB 1711 ATGCATTCATTTCTGG 1727

RESULT 5
US-09-737-698B-27/c
Sequence 27, Application US/09737698B

Patent No. 6462258
GENERAL INFORMATION:
APPLICANT: Fincher, Karen
APPLICANT: Wilkinson, Jack
TITLE OF INVENTION: No. 6462258el Plant Expression Constructs
FILE REFERENCE: 38-21(51499)C
CURRENT APPLICATION NUMBER: US/09/737,698B
CURRENT FILING DATE: 2000-12-15
PRIOR APPLICATION NUMBER: US 60/171,173
PRIOR FILING DATE: 1999-12-16
NUMBER OF SEQ ID NOS: 30
SOFTWARE: PatentIn version 3.0
SEQ ID NO 27
LENGTH: 1946
TYPE: DNA
ORGANISM: artificial sequence
FEATURE:
NAME/KEY: promoter
LOCATION: (1)..(1946)
OTHER INFORMATION: chimeric promoter fusion FMV and Act11 polynucleotides + Act1
US-09-737-698B-27

Query Match 4.2%; Score 17; DB 4; Length 1946;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 325 TCTGATGATTCATTT 341
|||||
DB 595 TCTGATGATTCATTT 579

RESULT 6
US-08-857-946-146
Sequence 146, Application US/08857946

Patent No. 5994075
GENERAL INFORMATION:
APPLICANT: Goodfellow, P.N.
TITLE OF INVENTION: METHODS FOR IDENTIFYING A MUTATION IN A
NUMBER OF SEQUENCES: 162
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Inc.
STREET: 75 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA

ZIP: 02109-1807
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/857,946
FILING DATE: 16-MAY-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/60/017,824
FILING DATE: 17-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kathleen M. Williams
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3529/05573
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111
INFORMATION FOR SEQ ID NO: 146:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: primer cktm300r
US-08-857-946-146

Query Match 4.0%; Score 16; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 214 CCATATTTCTTGAC 229
Db 17 CCATATTTCTTGAC 32

RESULT 7
US-08-970-740-146
Sequence 146, Application US/08970740
Patent No. 6015670
GENERAL INFORMATION:
APPLICANT: Goodfellow, P.N.
TITLE OF INVENTION: METHODS FOR IDENTIFYING A MUTATION IN A
TITLE OF INVENTION: GENE OF INTEREST
NUMBER OF SEQUENCES: 162
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Inc.
STREET: 28 State Street, 28th Floor
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/970,740
FILING DATE: 14-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/857,946
FILING DATE: 16-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/017,824
FILING DATE: 17-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Kathleen M. Williams
REGISTRATION NUMBER: 34,380

REFERENCE/DOCKET NUMBER: 3529/59829
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-7111
TELEFAX: 617-227-4399
INFORMATION FOR SEQ ID NO: 146:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: primer cktm300r
US-08-970-740-146

Query Match 4.0%; Score 16; DB 3; Length 40;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 214 CCATATTTCTTGAC 229
Db 17 CCATATTTCTTGAC 32

RESULT 8
US-07-918-953-3/C
Sequence 3, Application US/07918953
Patent No. 526453
GENERAL INFORMATION:
APPLICANT: ANDY, ROBIN J.
TITLE OF INVENTION: TISSUE-SELECTIVE INSULIN ANALOGS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/918,953
FILING DATE: 19920730
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: BOZICEVIC, KARL
REGISTRATION NUMBER: 28,807
REFERENCE/DOCKET NUMBER: 21900-2027420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-677-7000
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 99 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-07-918-953-3

Query Match 4.0%; Score 16; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 191 TCCAGGGCCAAAGGCT 206
Db 92 TCCAGGGCCAAAGGCT 77

RESULT 9
US-08-081-661-3/C
Sequence 3, Application US/08081661
Patent No. 5446020
GENERAL INFORMATION:
APPLICANT: ANDY, ROBIN J.
APPLICANT: LARSON, ERIC R.
TITLE OF INVENTION: TISSUE-SELECTIVE INSULIN ANALOGS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,661
FILING DATE: 23-JUN-1993
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/918,953
FILING DATE: 30-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: BOZICEVIC, KARL
REGISTRATION NUMBER: 28, 807
REFERENCE/DOCKET NUMBER: 21900-2027420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-677-7000
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 99 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-08-081-661-3
Query Match 4.0%; Score 16; DB 1; Length 99;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 191 TCCAGGCCAAGGCT 206
DB 92 TCCAGGCCAAGGCT 77
RESULT 10
US-08-487-001A-32
Sequence 32, Application US/08487001A
Patent No. 5795862
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELS, LYNDIA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA
PROTEINS AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.

ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,001A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Verter, Carol Talkington
REGISTRATION NUMBER: 37,459
REFERENCE/DOCKET NUMBER: 2618-17-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..242
US-08-487-001A-32
Query Match 4.0%; Score 16; DB 1; Length 242;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 87 AATTATGCTCAGATG 102
DB 129 AATTATGCTCAGATG 144
RESULT 11
US-08-630-822A-32
Sequence 32, Application US/08630822A
Patent No. 5840695
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELS, LYNDIA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,822A
FILING DATE: 11-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 32:

SEQUENCE CHARACTERISTICS:
LENGTH: 242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..242
US-08-630-822A-32

Query Match 4.0%; Score 16; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 AATTATGCTCAGATG 102
|||||
Db 129 AATTATGCTCAGATG 144

RESULT 12

US-09-005-069-32
Sequence 32, Application US/09005069
Patent No. 5932470
GENERAL INFORMATION:
APPLICANT: FRANK, GLENN R.
APPLICANT: HUNTER, SHIRLEY WU
APPLICANT: WALLENFELS, LYNDIA
TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,069
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,822
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: CONNELL, GARY J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-17-C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 3..242
US-09-005-069-32

Query Match 4.0%; Score 16; DB 2; Length 242;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 AATTATGCTCAGATG 102
|||||
Db 129 AATTATGCTCAGATG 144

RESULT 13

US-07-918-953-12/C
Sequence 12, Application US/07918953
Patent No. 5268453
GENERAL INFORMATION:
APPLICANT: ANDY, ROBIN J.
APPLICANT: LARSON, ERIC R.
TITLE OF INVENTION: TISSUE-SELECTIVE INSULIN ANALOGS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/918,953
FILING DATE: 19920730
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: BOZICEVIC, KARL
REGISTRATION NUMBER: 28,807
REFERENCE/DOCKET NUMBER: 21900-2027420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-677-7000
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 209..271
OTHER INFORMATION: /note="The Insulin analog
OTHER INFORMATION: is numbered from amino acid number 1 of the
OTHER INFORMATION: A chain."
US-07-918-953-12

Query Match 4.0%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 191 TCACAGGCCAAGGCT 206
|||||
Db 189 TCACAGGCCAAGGCT 174

RESULT 14
US-07-918-953-14/C
Sequence 14, Application US/07918953
Patent No. 5268453
GENERAL INFORMATION:
APPLICANT: ANDY, ROBIN J.
APPLICANT: LARSON, ERIC R.
TITLE OF INVENTION: TISSUE-SELECTIVE INSULIN ANALOGS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER

STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/918,953
FILING DATE: 19920730
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: BOZICEVIC, KARL
REGISTRATION NUMBER: 28,807
REFERENCE/DOCKET NUMBER: 21900-2027420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-677-7000
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 209..271
OTHER INFORMATION: /note="The insulin analog
OTHER INFORMATION: is numbered from amino acid number 1 of the
OTHER INFORMATION: A chain."
US-07-918-953-14

Query Match 4.0%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 TCCAGGGCCCAAGGCT 206
|||||
DB 189 TCCAGGGCCCAAGGCT 174

RESULT 15
US-08-081-661-12/c
Sequence 12, Application US/08081661
Patent No. 5446020
GENERAL INFORMATION:
APPLICANT: ANDY, ROBIN J.
APPLICANT: LARSON, ERIC R.
TITLE OF INVENTION: TISSUE-SELECTIVE INSULIN ANALOGS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,661
FILING DATE: 23-JUN-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/918,953

FILING DATE: 30-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: BOZICEVIC, KARL
REGISTRATION NUMBER: 28,807
REFERENCE/DOCKET NUMBER: 21900-2027420
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-677-7000
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 209..271
OTHER INFORMATION: /note="The insulin analog
OTHER INFORMATION: is numbered from amino acid number 1 of the
OTHER INFORMATION: A chain."
US-08-081-661-12

Query Match 4.0%; Score 16; DB 1; Length 275;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 191 TCCAGGGCCCAAGGCT 206
|||||
DB 189 TCCAGGGCCCAAGGCT 174

Search completed: January 2, 2003, 16:01:03
Job time : 78 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 16:01:07 : Search time 84 Seconds
(without alignments)
2069.842 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402

Sequence: 1 ccggcgccggcgtaccacgct.....ggcatactctctgttcc 402

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapept 60.0

Searched: 381593 seqs, 216252194 residues

Word size : 0

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	5.5	574	12	US-10-062-254-335 Sequence 335, App
2	18	4.5	872	10	US-09-925-297-3358 Sequence 358, App
3	18	4.5	2000	9	US-09-938-842A-3855 Sequence 3855, App
4	18	4.5	4623	10	US-09-864-864-288 Sequence 288, App
5	18	4.5	25277	10	US-09-764-878-285 Sequence 285, App
6	18	4.5	25279	10	US-09-764-860-1148 Sequence 1148, App
7	18	4.5	25279	10	US-09-764-878-284 Sequence 284, App
8	18	4.5	25279	10	US-09-764-860-1147 Sequence 1147, App
9	17	4.2	179	10	US-09-954-456-1496 Sequence 1496, App
10	17	4.2	179	10	US-09-880-107-2996 Sequence 2996, App
11	17	4.2	219	9	US-09-796-692-7632 Sequence 7632, App
12	17	4.2	260	10	US-09-878-574-11983 Sequence 11983, App
13	17	4.2	498	10	US-09-925-297-346 Sequence 346, App
14	17	4.2	1183	9	US-10-008-118A-33 Sequence 33, App
15	17	4.2	1183	10	US-09-443-704-33 Sequence 33, App
16	17	4.2	1193	10	US-09-737-626A-11 Sequence 11, App
17	17	4.2	1946	10	US-09-737-626A-27 Sequence 27, App
18	17	4.2	2982	9	US-09-938-842A-398 Sequence 398, App
19	17	4.2	3597	9	US-10-098-841-84 Sequence 84, App

ALIGNMENTS

20	16	4.0	94	10	US-09-864-761-30732	Sequence 30732, A
c 21	16	4.0	108	10	US-09-563-817-655	Sequence 655, App
c 22	16	4.0	138	10	US-09-974-300-5722	Sequence 5722, App
c 23	16	4.0	201	10	US-09-789-523-20	Sequence 20, App
c 24	16	4.0	228	10	US-09-867-701-7968	Sequence 7968, App
c 25	16	4.0	242	7	US-08-809-423A-32	Sequence 32, App
c 26	16	4.0	269	10	US-09-878-574-7917	Sequence 7917, App
c 27	16	4.0	320	10	US-09-899-917-14	Sequence 14, App
c 28	16	4.0	350	9	US-09-796-692-7386	Sequence 7386, App
c 29	16	4.0	351	10	US-09-280-030-49	Sequence 49, App
c 30	16	4.0	374	9	US-09-938-842A-4736	Sequence 4736, App
c 31	16	4.0	385	10	US-09-864-761-14179	Sequence 14179, A
c 32	16	4.0	390	10	US-09-280-030-48	Sequence 48, App
c 33	16	4.0	395	10	US-09-960-352-10581	Sequence 10581, A
c 34	16	4.0	420	10	US-09-864-761-10796	Sequence 10796, A
c 35	16	4.0	428	10	US-09-783-590-8704	Sequence 8704, App
c 36	16	4.0	450	9	US-09-804-409A-10	Sequence 10, App
c 37	16	4.0	457	10	US-09-899-917-12	Sequence 12, App
c 38	16	4.0	468	9	US-09-796-692-4911	Sequence 4911, App
c 39	16	4.0	468	9	US-09-796-692-9012	Sequence 9012, App
c 40	16	4.0	484	10	US-09-867-550-1781	Sequence 1781, App
c 41	16	4.0	512	9	US-09-728-444-427	Sequence 427, App
c 42	16	4.0	514	10	US-09-899-917-11	Sequence 11, App
c 43	16	4.0	531	10	US-09-864-761-6877	Sequence 6877, App
c 44	16	4.0	576	10	US-09-864-761-9132	Sequence 9132, App
c 45	16	4.0	581	10	US-09-864-761-9876	Sequence 9876, App

RESULT 1
US-10-062-254-335
Sequence 335, Application US/10062254
Patent No. US2002013882A1

GENERAL INFORMATION:

APPLICANT: Cahoon, Edgar B
APPLICANT: Falcio, Rebecca E
APPLICANT: Fang, Yiwen
APPLICANT: Hantke, Sabine S.
APPLICANT: Lee, Jian-Ming
APPLICANT: Li, Zhongsen
APPLICANT: Miao, Guo-Hua
APPLICANT: Morgante, Michele
APPLICANT: Niu, Xiping
APPLICANT: Odell, Joan
APPLICANT: Rafalski, Antoni
APPLICANT: Sakai, Hajime
APPLICANT: Zheng, Peizhong
APPLICANT: Zhu, Qun
TITLE OF INVENTION: Polynucleotides Encoding Proteins Involved in Plant Metabolism
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/062,254
CURRENT FILING DATE: 2002-02-01
PRIOR APPLICATION NUMBER: 09/630,346
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: 60/146511
PRIOR FILING DATE: 1999-07-30
PRIOR APPLICATION NUMBER: 60/156006
PRIOR FILING DATE: 1999-09-23
PRIOR APPLICATION NUMBER: 60/156899
PRIOR FILING DATE: 1999-09-30
PRIOR APPLICATION NUMBER: 60/157297
PRIOR FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: 60/169767
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/171054
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: 60/172958
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: 60/171515
PRIOR FILING DATE: 1999-12-22

```
; PRIOR APPLICATION NUMBER: 60/173535
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 375
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 335
; LENGTH: 574
; TYPE: DNA
; ORGANISM: glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (495)
; NAME/KEY: unsure
; LOCATION: (511)
; NAME/KEY: unsure
; LOCATION: (528)
; NAME/KEY: unsure
; LOCATION: (538)
; US-10-062-254-335
```

```
Query Match 5.5%; Score 22; DB 12; Length 574;
Best Local Similarity 100.0%; Pred. No. 0.046;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 131 GATGAGAGAGAGAGAGAT 152
DB 95 GATGAGAGAGAGAGAGAT 116
```

RESULT 2

```
US-09-925-297-358
; Sequence 358, Application US/09925297
; Patent No. US2002081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 358
; LENGTH: 872
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (803)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (813)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (871)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-925-297-358
```

```
Query Match 4.5%; Score 18; DB 10; Length 872;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 134 GAGAGAGAGAGAGAGA 151
DB 468 GAGAGAGAGAGAGAGA 485
```

RESULT 3

```
US-09-938-842A-3855/C
; Sequence 3855, Application US/09938842A
; Patent No. US20020160378A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SAME, AND METHODS OF USE
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO: 3855
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; US-09-938-842A-3855
```

```
Query Match 4.5%; Score 18; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 6.8;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 123 TAAGCTCGATGAGAGA 140
DB 1630 TAAGCTCGATGAGAGA 1613
```

RESULT 4

```
US-09-864-864-288
; Sequence 288, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, David C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: AND DIAGNOSIS OF OVARIAN CANCER
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Corlax Invention Disclosure Database
; SEQ ID NO: 288
; LENGTH: 4623
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-864-864-288
```

```
Query Match 4.5%; Score 18; DB 10; Length 4623;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 134 GAGAGAGAGAGAGAGA 151
DB 1072 GAGAGAGAGAGAGAGA 1089
```

RESULT 5

```
US-09-764-878-285/C
; Sequence 285, Application US/09764878
```

```

: Patent No. US20020090615A1
: GENERAL INFORMATION:
:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PA121
: CURRENT APPLICATION NUMBER: US/09/764, 878
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 428
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 285
: LENGTH: 25277
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-764-878-285

```

Query Match	4.5%;	Score 18;	DB 10;	Length 25277;
Best Local Similarity	100.0%;	Pred. NO. 8.4;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	241	AATTTTGAAGGCAATAT	258
Db	2821	AATTTTAGAAGGCAATAT	2804

```

RESULT 6
US-09-764-860-1148/c
: Sequence 1148, Application US/09764860
: Patent No. US20020094953A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC008
: CURRENT APPLICATION NUMBER: US/09/764,860
: PRIOR FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1198
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1148
: LENGTH: 25277
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-764-860-1148

```

Query Match	4.58;	Score 18;	DB 10;	Length 25277;
Best Local Similarity	100.0%;	Pred. No. 8.4;		
Matches 18;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	241	AATTTTGAAGGCAATAT	258
Db	2821	AATTTTGAAGGCAATAT	2804

```

: RESULT 7
: US-09-764-878-284/c
: Sequence 284. Application US/09764878
: Patent No. US20020090615A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PA121
: CURRENT APPLICATION NUMBER: US/09/764,878
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 428
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 284
: LENGTH: 25279
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-764-878-284

```

Query Match	4.5%	Score 18;	DB 10;	Length 25279;
-------------	------	-----------	--------	---------------

```
Best Local Similarity    100.0%; Pred. No. 8.4;
Matches      18; Conservative      0; Mismatches      0; Indels      0; Gaps      0.
```

RESULT 8
US-09-76

US-09-764-860-1147/c
; Sequence 1147, Application US/09764860
; Patent No. US20020094953A1

```

1  APPLICANT:  Kosen et al.
2  TITLE OF INVENTION:  Nucleic Acids, Proteins, and Antibodies
3  FILE REFERENCE:  PC008
4  CURRENT APPLICATION NUMBER:  US/09/764,860
5  CURRENT FILING DATE:  2001-01-17
6  Prior application date removed - consult PALM or file wrappr
7  NUMBER OF SEQ ID NOS:  1198
8  SOFTWARE:  PatentIn Ver. 2.0
9  SEQ ID NO 1147
10 LENGTH: 25279
11 TYPE: DNA
12 ORGANISM: Homo sapiens
13 US-09-764-860-1147

```

Query Match	4.5%	Score 18:	DB 10:	Length 25279:
Best Local Similarity	100.0%	Pred. No. 8.4:		
Matches 18:	Conservative 0:	Mismatches 0:	Indels 0:	Gaps 0

Qy	241	AATTTTGAAGGCAATAT	258
Db	2821	AATTTTGAAGGCAATAT	2804

```

RESULT 9
US-09-954-456-1496
; Sequence 1496, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C
; TITLE OF INVENTION: Sets
; FILE REFERENCE: 689290-76
; CURRENT APPLICATION NUMBER: US/09/954,456
; CURRENT FILING DATE: 2001-09-18
; PRIOR APPLICATION NUMBER: US/60/233,617
; PRIOR FILING DATE: 2000-09-18
; PRIOR APPLICATION NUMBER: US/60/234,052
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US/60/234,923
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,134
; PRIOR FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: US/60/235,637
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,638
; PRIOR FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US/60/235,711
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,720
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,840
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: US/60/235,863
; PRIOR FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 2276
; SOFTWARE: patentIn version 3.0
; SEQ ID NO 1496
; LENGTH: 179
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: n=a,t,g or c
US-09-954-456-1496
```

```
Query Match          4.2%; Score 17; DB 10; Length 179;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 135 AGAAGAGGAAGAGAGA 151
    ||||||||||||||||
DB 71 AGAAGAGGAAGAGAGA 87
```

```
RESULT 10
US-09-880-107-2996
; Sequence 2996, Application US/09880107
; Patent No. US20020142981A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
```

```
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
```

```
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
```

```
; FILE REFERENCE: 44921-5028-WO
```

```
; CURRENT APPLICATION NUMBER: US/09/880,107
```

```
; CURRENT FILING DATE: 2001-06-14
```

```
; PRIOR APPLICATION NUMBER: US 60/211,379
```

```
; PRIOR FILING DATE: 2000-06-14
```

```
; PRIOR APPLICATION NUMBER: US 60/237,054
```

```
; PRIOR FILING DATE: 2000-10-02
```

```
; NUMBER OF SEQ ID NOS: 3950
```

```
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2996
```

```
; LENGTH: 179
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 R95966
```

```
; NAME/KEY: unsure
```

```
; LOCATION: (1)..(179)
```

```
; OTHER INFORMATION: n = a or c or g or t
```

```
US-09-880-107-2996
```

```
Query Match          4.2%; Score 17; DB 10; Length 179;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 135 AGAAGAGGAAGAGAGA 151
    ||||||||||||||||
DB 71 AGAAGAGGAAGAGAGA 87
```

```
RESULT 11
```

```
US-09-796-692-7632/c
```

```
; Sequence 7632, Application US/09796692
```

```
; Publication No. US20020198362A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Gaiger, Alexander
```

```
; APPLICANT: Algate, Paul A.
```

```
; APPLICANT: Mannion, Jane
```

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
```

```
; FILE REFERENCE: 2077,001200
```

```
; CURRENT APPLICATION NUMBER: US/09/796,692
```

```
; CURRENT FILING DATE: 2001-03-01
```

```
; PRIOR APPLICATION NUMBER: 60/186,126
```

```
; PRIOR FILING DATE: 2000-03-01
```

```
; PRIOR APPLICATION NUMBER: 60/190,479
```

```
; PRIOR FILING DATE: 2000-03-17
```

```
; PRIOR APPLICATION NUMBER: 60/200,545
```

```
; PRIOR FILING DATE: 2000-04-27
```

```
; PRIOR APPLICATION NUMBER: 60/200,303
```

```
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
```

```
; PRIOR APPLICATION NUMBER: 60/202,084
```

```
; PRIOR FILING DATE: 2000-05-04
```

```
; PRIOR APPLICATION NUMBER: 60/206,201
```

```
; PRIOR FILING DATE: 2000-05-22
```

```
; PRIOR APPLICATION NUMBER: 60/218,950
```

```
; PRIOR FILING DATE: 2000-07-14
```

```
; PRIOR APPLICATION NUMBER: 60/222,903
```

```
; PRIOR FILING DATE: 2000-08-03
```

```
; PRIOR APPLICATION NUMBER: 60/223,416
```

```
; PRIOR FILING DATE: 2000-08-04
```

```
; PRIOR APPLICATION NUMBER: 60/223,378
```

```
; PRIOR FILING DATE: 2000-08-07
```

```
; NUMBER OF SEQ ID NOS: 9597
```

```
; SOFTWARE: FastSeq for Windows Version 3.0
```

```
; SEQ ID NO 7632
```

```
; LENGTH: 219
```

```
; TYPE: DNA
```

```
; ORGANISM: Homo sapiens
```

```
; FEATURE:
```

```
; NAME/KEY: unsure
```

```
; LOCATION: (151)
```

```
; OTHER INFORMATION: n=A,T,C or G
```

```
; NAME/KEY: unsure
```

```
; LOCATION: (187)
```

```
; OTHER INFORMATION: n=A,T,C or G
```

```
; NAME/KEY: unsure
```

```
; LOCATION: (211)
```

```
; OTHER INFORMATION: n=A,T,C or G
```

```
; NAME/KEY: unsure
```

```
; LOCATION: (216)
```

```
; OTHER INFORMATION: n=A,T,C or G
```

```
US-09-796-692-7632
```

```
Query Match          4.2%; Score 17; DB 9; Length 219;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 136 GAGAAGGAAGAGAGAT 152
    ||||||||||||||||
DB 176 GAGAAGGAAGAGAGAT 160
```

```
RESULT 12
US-09-878-574-11983/c
```

```
; Sequence 11983, Application US/09878574
```

```
; Patent No. US20020110548A1
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Byrum, Joseph R.
```

```
; APPLICANT: La Rosa, Thomas J.
```

```
; APPLICANT: Thompson, Michael D.
```

```
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
```

```
; FILE REFERENCE: 38-21(15401)B
```

```
; CURRENT APPLICATION NUMBER: US/09/878,574
```

```
; CURRENT FILING DATE: 2001-12-21
```

```
; PRIOR APPLICATION NUMBER: 09/333,535
```

```
; PRIOR FILING DATE: 1999-06-14
```

```
; NUMBER OF SEQ ID NOS: 15775
```

```
; SEQ ID NO 11983
```

```
; LENGTH: 260
```

```
; TYPE: DNA
```

```
; ORGANISM: Glycine max
```

```
; OTHER INFORMATION: Clone ID: 701065166H1
```

```
US-09-878-574-11983
```

```
Query Match          4.2%; Score 17; DB 10; Length 260;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


QY 353 TGAACACAGTACAAA 369
 ||||||||||||||||
 Db 37 TGAACACAGTACAAA 21

RESULT 13
 US-09-925-297-346/c
 ; Sequence 346, Application US/09925297
 ; Patent No. US20020081659A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
 ; FILE REFERENCE: PA105
 ; CURRENT APPLICATION NUMBER: US/09/925,297
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: PCT/US00/05989
 ; PRIOR FILING DATE: 2000-03-08
 ; PRIOR APPLICATION NUMBER: 60/124,270
 ; PRIOR FILING DATE: 1999-03-12
 ; NUMBER OF SEQ ID NOS: 928
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 346
 ; LENGTH: 498
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (252)
 ; OTHER INFORMATION: n equals a,t,g, or c
 ; NAME/KEY: misc feature
 ; LOCATION: (493)
 ; OTHER INFORMATION: n equals a,t,g, or c.
 US-09-925-297-346

Query Match 4.2%; Score 17; DB 10; Length 498;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 TTCCAGGGCCAGGGCT 206
 ||||||||||||||||
 Db 300 TTCCAGGGCCAGGGCT 284

RESULT 14
 US-10-008-118A-33/c
 ; Sequence 33, Application US/10008118A
 ; Publication No. US20020187539A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cahoon, Rebecca E.
 ; TITLE OF INVENTION: Plant MYB-Related Transcription Factors
 ; FILE REFERENCE: B01280 USDIY
 ; CURRENT APPLICATION NUMBER: US/10/008,118A
 ; CURRENT FILING DATE: 2001-12-05
 ; PRIOR APPLICATION NUMBER: 60/109,294
 ; PRIOR FILING DATE: 1998-11-20
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 33
 ; LENGTH: 1183
 ; TYPE: DNA
 ; ORGANISM: Triticum aestivum
 US-10-008-118A-33

Query Match 4.2%; Score 17; DB 9; Length 1183;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 AAGAAGAAAGAGATG 153
 ||||||||||||||||
 Db 1089 AAGAAGAAAGAGATG 1073

RESULT 15
 US-09-443-704-33/c
 ; Sequence 33, Application US/09443704
 ; Patent No. US20020066120A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cahoon, Rebecca E.
 ; APPLICANT: Liu, Zhan-Bin
 ; APPLICANT: Odeh, Joan
 ; APPLICANT: Rafalski, Antoni
 ; APPLICANT: Shi, June
 ; APPLICANT: Weng, Zude
 ; TITLE OF INVENTION: Plant MYB-Related Transcription Factors
 ; FILE REFERENCE: B01280 US NA
 ; CURRENT APPLICATION NUMBER: US/09/443,704
 ; CURRENT FILING DATE: 1999-11-19
 ; EARLIER APPLICATION NUMBER: 60/109,294
 ; EARLIER FILING DATE: No. US20020066120A1ember 20, 1998
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: Microsoft Office 97
 ; SEQ ID NO 33
 ; LENGTH: 1183
 ; TYPE: DNA
 ; ORGANISM: Triticum aestivum
 US-09-443-704-33

Query Match 4.2%; Score 17; DB 10; Length 1183;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 AAGAAGAAAGAGATG 153
 ||||||||||||||||
 Db 1089 AAGAAGAAAGAGATG 1073

Search completed: January 2, 2003, 17:49:29
 Job time : 134 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 16:07:27 : Search time 147 Seconds
(without alignments)
2271.753 Million cell updates/sec

Title: US-09-850-147-1

Sequence: 1 ccggcgccggcgctcaccagct.....ggcatacttcctgtgtcc 402

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 655294 seqs, 415357811 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1310588

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq.*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	4.5	19	US-10-310-188-10261	Sequence 10261, A
2	18	4.5	319608	US-09-539-333D-1	GENERAL INFORMATION
3	18	4.5	319608	US-09-539-333D-1	Sequence 1, Appl1
4	17	4.2	236	US-09-534-850-21604	Sequence 21604, A
5	17	4.2	238	US-09-534-850-20438	Sequence 20438, A
6	17	4.2	234	US-09-513-999C-20190	Sequence 20190, A
7	17	4.2	261	US-09-513-999C-14719	Sequence 14719, A
8	17	4.2	324	US-09-513-999C-33198	Sequence 33198, A
9	17	4.2	489	US-10-092-411A-66	Sequence 66, Appl
10	17	4.2	597	US-10-092-411A-1790	Sequence 1790, Ap
11	17	4.2	1165	US-09-724-676-25002	Sequence 25002, A
12	17	4.2	1165	US-09-724-676-25002	Sequence 25002, A
13	17	4.2	1199	US-09-724-676-25006	Sequence 25006, A
14	17	4.2	1199	US-09-724-676-25006	Sequence 25006, A
15	17	4.2	1203	US-09-724-676-25004	Sequence 25004, A
16	17	4.2	1203	US-09-724-676-25004	Sequence 25004, A
17	17	4.2	1237	US-09-724-676-25008	Sequence 25008, A
18	17	4.2	1237	US-09-724-676-25008	Sequence 25008, A
19	17	4.2	1285	US-09-724-676-25003	Sequence 25003, A
20	17	4.2	1285	US-09-724-676-25003	Sequence 25003, A
21	17	4.2	1319	US-09-724-676-25007	Sequence 25007, A
22	17	4.2	1319	US-09-724-676-25007	Sequence 25007, A
23	17	4.2	1355	US-09-724-676-25005	Sequence 25005, A
24	17	4.2	1355	US-09-724-676-25005	Sequence 25005, A
25	17	4.2	1389	US-09-724-676-25009	Sequence 25009, A
26	17	4.2	1389	US-09-724-676-25009	Sequence 25009, A

c 27	17	4.2	1536	5	US-09-724-676-24992	Sequence 24992, A
c 28	17	4.2	1536	5	US-09-724-676-24992	Sequence 24992, A
c 29	17	4.2	1574	5	US-09-724-676-24994	Sequence 24994, A
c 30	17	4.2	1574	5	US-09-724-676-24994	Sequence 24994, A
c 31	17	4.2	1595	5	US-09-724-676-24987	Sequence 24987, A
c 32	17	4.2	1595	5	US-09-724-676-24987	Sequence 24987, A
c 33	17	4.2	1633	5	US-09-724-676-24989	Sequence 24989, A
c 34	17	4.2	1633	5	US-09-724-676-24989	Sequence 24989, A
c 35	17	4.2	1656	5	US-09-724-676-24993	Sequence 24993, A
c 36	17	4.2	1656	5	US-09-724-676-24993	Sequence 24993, A
c 37	17	4.2	1715	5	US-09-724-676-24988	Sequence 24988, A
c 38	17	4.2	1715	5	US-09-724-676-24988	Sequence 24988, A
c 39	17	4.2	1726	5	US-09-724-676-24995	Sequence 24995, A
c 40	17	4.2	1726	5	US-09-724-676-24995	Sequence 24995, A
c 41	17	4.2	1785	5	US-09-724-676-24990	Sequence 24990, A
c 42	17	4.2	1785	5	US-09-724-676-24990	Sequence 24990, A
c 43	17	4.2	1986	5	US-09-724-676-23591	Sequence 23591, A
c 44	17	4.2	1986	5	US-09-724-676-23591	Sequence 23591, A
c 45	17	4.2	263744	6	US-10-229-834A-6	Sequence 6, Appl1

ALIGNMENTS

```

RESULT 1
US-10-310-188-10261/c
; Sequence 10261, Application US/10310188
; GENERAL INFORMATION:
; APPLICANT: RosettaGenomics
; TITLE OF INVENTION: BIOINFORMATIALLY DETECTABLE GROUP OF NOVEL VIRAL REGULATORY G
; FILE REFERENCE: 47487
; CURRENT APPLICATION NUMBER: US/10/310,188
; CURRENT FILING DATE: 2002-12-19
; NUMBER OF SEQ ID NOS: 86841
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10261
; LENGTH: 19
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-310-188-10261

Query Match      4.5%  Score 18; DB 6; Length 19;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      134 GAGAGAAGAGAAGAGA 151
Db      19 GAGAGAAGAGAAGAGA 2

RESULT 2
US-09-539-333D-1/c
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essiox, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENST.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,303
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915

```

```
, PRIOR FILING DATE: 1999-07-27
, PRIOR APPLICATION NUMBER: US 60/146,453
, PRIOR FILING DATE: 1999-07-29
, PRIOR APPLICATION NUMBER: US 60/146,452
, PRIOR FILING DATE: 1999-07-29
, PRIOR APPLICATION NUMBER: US 60/162,288
, PRIOR FILING DATE: 1999-10-28
, PRIOR APPLICATION NUMBER: US 09/416,384
, PRIOR FILING DATE: 1999-10-12
, NUMBER OF SEQ ID NOS: 231
, SOFTWARE: Patent.pm
, FEATURE:
, NAME/KEY: allele
, LOCATION: 110222
, OTHER INFORMATION: polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 111978
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 112468
, OTHER INFORMATION: polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 117324..117327
, OTHER INFORMATION: deletion ACTT
, FEATURE:
, NAME/KEY: allele
, LOCATION: 118972
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 119160..119161
, OTHER INFORMATION: deletion TT
, FEATURE:
, NAME/KEY: allele
, LOCATION: 119316
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 119321
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 119526
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 120573
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 121527
, OTHER INFORMATION: polymorphic base A or C
, FEATURE:
, NAME/KEY: allele
, LOCATION: 126105
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 129789
, OTHER INFORMATION: polymorphic base C or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 130777
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 136942..136944
, OTHER INFORMATION: deletion ATT
, FEATURE:
, NAME/KEY: allele
, LOCATION: 143839
, OTHER INFORMATION: polymorphic base A or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 146668
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 147281
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 147505
, OTHER INFORMATION: polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 148372
, OTHER INFORMATION: polymorphic base A or C
, FEATURE:
, NAME/KEY: allele
, LOCATION: 149012
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 149113
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 151637
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 151748
, OTHER INFORMATION: deletion G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 151769
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 151847
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 152691
, OTHER INFORMATION: polymorphic base A or C
, FEATURE:
, NAME/KEY: allele
, LOCATION: 152766
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 153046
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 153123
, OTHER INFORMATION: polymorphic base A or G
, FEATURE:
, NAME/KEY: allele
, LOCATION: 153925
, OTHER INFORMATION: polymorphic base C or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 153977
, OTHER INFORMATION: polymorphic base G or T
, FEATURE:
, NAME/KEY: allele
, LOCATION: 154502
```

OTHER INFORMATION: polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 154677
OTHER INFORMATION: polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 154879
OTHER INFORMATION: polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 155802
OTHER INFORMATION: polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 156448
OTHER INFORMATION: polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 157238
OTHER INFORMATION: polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 157897
OTHER INFORMATION: polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 158172
OTHER INFORMATION: polymorphic base A or G
FEATURE:

Query Match 4.5%; Score 18; DB 5; Length 319608;
Best local similarity 100.0%; Pred. No. 9.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 134 GAGAGAAGGAAGAGA 151
Db 75504 GAGAGAAGGAAGAGA 75487

RESULT 3
US-09-539-333D-1/c
Sequence 1, Application US/09539333D
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Bihain, Bernard
APPLICANT: Essioux, Laurent
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
FILE REFERENCE: GENSET 047A1S
CURRENT APPLICATION NUMBER: US/09/539, 333D
CURRENT FILING DATE: 2000-03-30
PRIOR APPLICATION NUMBER: US 60/126,903
PRIOR FILING DATE: 1999-03-30
PRIOR APPLICATION NUMBER: US 60/131,971
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/132,065
PRIOR FILING DATE: 1999-04-30
PRIOR APPLICATION NUMBER: US 60/143,928
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 60/145,915
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: US 60/146,453
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/146,452
PRIOR FILING DATE: 1999-07-29
PRIOR APPLICATION NUMBER: US 60/162,288

PRIOR FILING DATE: 1999-10-28
PRIOR APPLICATION NUMBER: US 09/416,384
PRIOR FILING DATE: 1999-10-12
NUMBER OF SEQ ID NOS: 231
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 31..1107
OTHER INFORMATION: 5'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 1108..1289
OTHER INFORMATION: exon A g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 14877..14920
OTHER INFORMATION: exon B g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 18778..18862
OTHER INFORMATION: exon Bbis g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 25593..25740
OTHER INFORMATION: exon C g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29388..29502
OTHER INFORMATION: exon D g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 29967..30282
OTHER INFORMATION: exon E g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 64666..64812
OTHER INFORMATION: exon F g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 65505..65853
OTHER INFORMATION: exon G g35018 gene
FEATURE:
NAME/KEY: misc_feature
LOCATION: 65854..67854
OTHER INFORMATION: 3'regulatory region g35018 gene
FEATURE:
NAME/KEY: exon
LOCATION: 94124..94964
OTHER INFORMATION: exon g35017
FEATURE:
NAME/KEY: exon
LOCATION: 201188..201234
OTHER INFORMATION: exon S g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U g35030 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216836..216915
OTHER INFORMATION: exon V g35030 gene
FEATURE:
NAME/KEY: misc_feature
LOCATION: 213818..215818
OTHER INFORMATION: 3'regulatory region g34872 gene

FEATURE:
NAME/KEY: exon
LOCATION: 215819..215941
OTHER INFORMATION: exon R complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 215819..215975
OTHER INFORMATION: exon Rbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 216661..216952
OTHER INFORMATION: exon Qbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 217061..217061
OTHER INFORMATION: exon Q complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 217061..217061
OTHER INFORMATION: exon Q1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 229647..229742
OTHER INFORMATION: exon X complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 230408..230721
OTHER INFORMATION: exon P complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231272..231412
OTHER INFORMATION: exon Obis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231787..231880
OTHER INFORMATION: exon O2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 231870..231879
OTHER INFORMATION: exon O1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 234174..234321
OTHER INFORMATION: exon O complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 237406..237428
OTHER INFORMATION: exon Nbis complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239807
OTHER INFORMATION: exon N2 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 239719..239853
OTHER INFORMATION: exon N complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240569
OTHER INFORMATION: exon M1117 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240596
OTHER INFORMATION: exon M1090 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240617
OTHER INFORMATION: exon M1069 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240644
OTHER INFORMATION: exon MS2 complement g34872 gene
FEATURE:

NAME/KEY: exon
LOCATION: 240528..240824
OTHER INFORMATION: exon M862 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..240994
OTHER INFORMATION: exon M692 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240528..241685
OTHER INFORMATION: exon M1 complement g34872 gene
FEATURE:
NAME/KEY: exon
LOCATION: 240800..240993
OTHER INFORMATION: exon MS1 complement g34872 gene
FEATURE:
NAME/KEY: misc.feature
LOCATION: 241686..243685
OTHER INFORMATION: 5' regulatory region g34872 gene
FEATURE:
NAME/KEY: misc.feature
LOCATION: 290652..292652
OTHER INFORMATION: 3' regulatory region g34665 gene
FEATURE:
NAME/KEY: exon
LOCATION: 292653..292841
OTHER INFORMATION: exon B complement g34665 gene

Query Match 4.5%: Score 18: DB 5: Length 319608;
Best Local Similarity 100.0%: Pred. No. 9.7;
Matches 18: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

Qy 134 GAGAGAGAGAGAGAGA 151
Db 75504 GAGAGAGAGAGAGAGA 75487

RESULT 4
US-09-534-850-21604
Sequence 21604, Application US/09534850
GENERAL INFORMATION:
APPLICANT: Seilhamer, Jeffrey J.
APPLICANT: Delegeane, Angelo M.
APPLICANT: Stuart, Susan G.
APPLICANT: Stuve, Laura L.
APPLICANT: Mullahy, Sara J.
APPLICANT: Naughton, Rebecca E.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING METABOLIC MOLECULES
FILE REFERENCE: PD-1020 CIP
CURRENT APPLICATION NUMBER: US/09/534, 850
CURRENT FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: 07/916,491
PRIOR FILING DATE: 1992-07-17
PRIOR APPLICATION NUMBER: 07/977,780
PRIOR FILING DATE: 1992-11-19
PRIOR APPLICATION NUMBER: 08/100,523
PRIOR FILING DATE: 1993-08-03
PRIOR APPLICATION NUMBER: 09/008,119
PRIOR FILING DATE: 1998-01-16
PRIOR APPLICATION NUMBER: 08/196,364
PRIOR FILING DATE: 1994-02-14
PRIOR APPLICATION NUMBER: 08/282,991
PRIOR FILING DATE: 1994-07-28
PRIOR APPLICATION NUMBER: 08/438,571
PRIOR FILING DATE: 1995-05-10
PRIOR APPLICATION NUMBER: 08/179,873
PRIOR FILING DATE: 1994-01-11
PRIOR APPLICATION NUMBER: 08/504,732
PRIOR FILING DATE: 1995-07-20
PRIOR APPLICATION NUMBER: 08/137,951
PRIOR FILING DATE: 1993-10-14
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 26730

```

; SOFTWARE: PERL Program
; SEQ ID NO 21604
; LENGTH: 236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00952490
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 210, 228
; OTHER INFORMATION: a, t, c, g, or other
US-09-534-850-21604

Query Match          4.28; Score 17; DB 5; Length 236;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      138 AGAAGGAAGGAGATGT 154
      |||
Db       125 AGAAGGAAGGAGATGT 141
```

```

RESULT 5
US-09-534-850-20438
; Sequence 20438, Application US/09534850
; GENERAL INFORMATION:
; APPLICANT: Sellhammer, Jeffrey J.
; APPLICANT: Delegeane, Angelo M.
; APPLICANT: Stuart, Susan G.
; APPLICANT: Stuve, Laura L.
; APPLICANT: Mullahy, Sara J.
; APPLICANT: Naughton, Rebecca E.
; TITLE OR INVENTION: POLYNUCLEOTIDES ENCODING OR REGULATING METABOLIC MOLECULES
; FILE REFERENCE: PD-1020 CIP
; CURRENT FILING DATE: 2000-03-24
; PRIOR FILING DATE: 1992-07-17
; PRIOR APPLICATION NUMBER: 07/916,491
; PRIOR FILING DATE: 1992-07-17
; PRIOR APPLICATION NUMBER: 07/977,780
; PRIOR FILING DATE: 1992-11-19
; PRIOR APPLICATION NUMBER: 08/100,523
; PRIOR FILING DATE: 1993-08-03
; PRIOR APPLICATION NUMBER: 09/008,119
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/196,364
; PRIOR FILING DATE: 1994-02-14
; PRIOR APPLICATION NUMBER: 08/282,991
; PRIOR FILING DATE: 1994-07-28
; PRIOR APPLICATION NUMBER: 08/438,571
; PRIOR FILING DATE: 1995-05-10
; PRIOR APPLICATION NUMBER: 08/179,873
; PRIOR FILING DATE: 1994-01-11
; PRIOR APPLICATION NUMBER: 08/504,732
; PRIOR FILING DATE: 1995-07-20
; PRIOR APPLICATION NUMBER: 08/137,951
; PRIOR FILING DATE: 1993-10-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 26730
; SOFTWARE: PERL Program
; SEQ ID NO 20438
; LENGTH: 238
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: hu00483915
US-09-534-850-20438
```

```

Query Match          4.28; Score 17; DB 5; Length 238;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      138 AGAAGGAAGGAGATGT 154
      |||
Db       89 AGAAGGAAGGAGATGT 105
```

```

RESULT 6
US-09-513-999C-20190/c
; Sequence 20190, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OR INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 20190
; LENGTH: 254
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 133
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 136
; OTHER INFORMATION: y=c or t
US-09-513-999C-20190
```

```

Query Match          4.28; Score 17; DB 5; Length 254;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      135 AGAAGGAAGGAAGAGA 151
      |||
Db       182 AGAAGGAAGGAAGAGA 166
```

```

RESULT 7
US-09-513-999C-14719/c
; Sequence 14719, Application US/09513999C
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OR INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14719
; LENGTH: 261
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 254
; OTHER INFORMATION: r=a or g
US-09-513-999C-14719
```

```

Query Match          4.28; Score 17; DB 5; Length 261;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      135 AGAAGGAAGGAAGAGA 151
```

DB 171 AGAAGAGGAAGAGA 155

RESULT 8

US-09-513-999C-33198

Sequence 33198, Application US/09513999C
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Duclet, A.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
FILE REFERENCE: 59.052.Reg
CURRENT APPLICATION NUMBER: US/09/513.999C
CURRENT FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/122,487
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 36681
SOFTWARE: Patent.pm
SEQ ID NO 33198
LENGTH: 324
TYPE: DNA
ORGANISM: Homo sapiens
US-09-513-999C-33198

Query Match 4.2%; Score 17; DB 5; Length 324;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 GCTCACACAGCTGCCCA 26
DB 121 GCTCACACAGCTGCCCA 137

RESULT 9

US-10-092-411A-66

Sequence 66, Application US/10092411A
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-101
CURRENT APPLICATION NUMBER: US/10/092.411A
CURRENT FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: US 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5676
SEQ ID NO 66
LENGTH: 489
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-10-092-411A-66

Query Match 4.2%; Score 17; DB 6; Length 489;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 223 CTTTGACTCAGGTGA 239
DB 332 CTTTGACTCAGGTGA 348

RESULT 10

US-10-092-411A-1790

Sequence 1790, Application US/10092411A
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-101
CURRENT APPLICATION NUMBER: US/10/092.411A
CURRENT FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: US 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5676
SEQ ID NO 1790
LENGTH: 597
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-10-092-411A-1790

Query Match 4.2%; Score 17; DB 6; Length 597;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 250 AGCAATATTAAATCT 266
DB 503 AGCAATATTAAATCT 519

RESULT 11

US-09-724-676-25002/C

Sequence 25002, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Comugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Comugen
CURRENT APPLICATION NUMBER: US/09/724.676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 25002
LENGTH: 1165
TYPE: DNA
ORGANISM: Homo sapiens
US-09-724-676-25002

Query Match 4.2%; Score 17; DB 5; Length 1165;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 AGAAGAGGAAGAGA 151
DB 218 AGAAGAGGAAGAGA 202

RESULT 12

US-09-724-676A-25002/C

Sequence 25002, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: Comugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Comugen
CURRENT APPLICATION NUMBER: US/09/724.676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 25002
LENGTH: 1165
TYPE: DNA
ORGANISM: Homo sapiens
US-09-724-676A-25002

Query Match 4.2%; Score 17; DB 5; Length 1165;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 135 AGAAGAGGAAGAGA 151

Db 218 AGAAGAAGAAAGGAGA 202

RESULT 13

US-09-724-676-25006/c
; Sequence 25006, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 2000-11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25006
; LENGTH: 1199
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-25006

Query Match

Best Local Similarity 4.2%; Score 17; DB 5; Length 1199;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 AGAAGAAGAAAGGAGA 151

Db 252 AGAAGAAGAAAGGAGA 236

RESULT 14

US-09-724-676A-25006/c
; Sequence 25006, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25006
; LENGTH: 1199
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-25006

Query Match

Best Local Similarity 4.2%; Score 17; DB 5; Length 1199;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 AGAAGAAGAAAGGAGA 151

Db 252 AGAAGAAGAAAGGAGA 236

RESULT 15

US-09-724-676-25004/c
; Sequence 25004, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 2000-11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25004
; LENGTH: 1203
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676-25004

Query Match 4.2%; Score 17; DB 5; Length 1203;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 135 AGAAGAAGAAAGGAGA 151
Db 218 AGAAGAAGAAAGGAGA 202

Search completed: January 2, 2003, 17:55:58
Job time : 426 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 15:04:27 ; Search time 2202 Seconds
(without alignments)
2956.669 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402
Sequence: 1 ccggcgccggcgcaccacgct.....gycataacttcctgtgttc 402

Scoring table: OLIGO_NMC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST: *
1: em_estbda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	7.2	352	12	BG354266 947033G03
2	29	7.2	365	13	B1388689 949044B01
3	29	7.2	400	13	B1396083 949044B01
4	29	7.2	479	10	AW091358 614094A10
5	29	7.2	487	10	BE598350 P11_81_E1
6	29	7.2	493	12	BG558648 RH122_58_

7	29	7.2	535	10	AW679158 WS1_22_A0
8	29	7.2	537	12	BG842699 MEST39-A0
9	29	7.2	568	10	AW565695 LG1_348_A
10	29	7.2	570	9	A1978092 614022D04
11	29	7.2	590	10	AW288508 618068C02
12	29	7.2	592	10	AW036911 614022D04
13	29	7.2	606	14	BO656076 PR0407_S
14	29	7.2	612	10	AW745589 WS1_35_A0
15	29	7.2	613	9	A1691590 606022D02
16	29	7.2	614	10	BE597103 P11_70_G0
17	29	7.2	681	13	BG946826 TP1_1_B10
18	29	7.2	1349	11	AY109457 Zea mays
19	29	5.5	356	9	A1619388 BE020893 sm53f03.Y
20	22	5.5	574	10	BE020893 sm53f03.Y
21	20	5.0	284	10	BB007187 BE020893 sm53f03.Y
22	20	5.0	300	10	BE401569 BM448247 DSA023D02
23	20	5.0	409	13	BM448247 DSA023D02
24	20	5.0	502	17	AO821035 HS_4728_B
25	20	5.0	696	17	AO781302 HS_3104_B
26	20	5.0	867	17	CNS055CG AL321865 Tetradon
27	20	5.0	928	14	BO685675 AGENCOURT
28	19	4.7	108	9	A1891239 A1891239 614020F10
29	19	4.7	315	17	TA122H01P AL462669 T. brucei
30	19	4.7	329	10	AW946083 AA734352 vv24f11.x
31	19	4.7	338	9	AA734352 vv24f11.x
32	19	4.7	343	17	BH849501 BF358635 OVO-ET000
33	19	4.7	347	12	BF358635 OVO-ET000
34	19	4.7	349	10	AW946080 AV031208
35	19	4.7	354	9	AV031208
36	19	4.7	357	10	AW755733 s108d10.Y
37	19	4.7	358	12	BF358608 OVO-ET000
38	19	4.7	363	12	BF358617 OVO-ET000
39	19	4.7	366	12	BF358620 OVO-ET000
40	19	4.7	372	17	BH104961 RCT-24-2
41	19	4.7	378	17	AZ638587 IM0498L05
42	19	4.7	413	10	AW924010 WS1_32_C0
43	19	4.7	414	14	BO454260 sa078d08.
44	19	4.7	416	17	AO280819 CTTB-E1-
45	19	4.7	423	10	AW704751 SK55a07.Y

ALIGNMENTS

RESULT 1
BG354266/c 352 bp mRNA linear EST 05-MAR-2001
DEFINITION 947033G03.xl 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
LOCUS
ACCESSION BG354266
VERSION BG354266.1 GI:13198328
KEYWORDS
SOURCE
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade: Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 352)
REFERENCE
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 947033 row: G column: 03.
Location/Qualifiers
1..352
/organism="Zea mays"

/cultivar="B73"
/db_xref="taxon:4577"
/clone_lib="947 - 2 week shoot from Barkan lab"
/tissue_type="leaf and stem, including leaf base"
/dev_stage="2 week old seedling (3 leaves)"
/lab_host="XLI-Blue"
/note="Organ: shoot; Vector: Lambda ZAP (pBluescript SK-);
Site_1: EcoRI; Site_2: XhoI; Directionally cloned using
Stratagene's Unitap XR cDNA cloning kit with the 5' end
at the EcoRI site. The library represents 8 x 10⁵
independent recombinant phage. The plants were greenhouse
grown."

BASE COUNT 106 a 75 c 65 g 106 t

ORIGIN

Query Match 7.2%; Score 29; DB 12; Length 352;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTCTATATTAACTTCGATGAGAGA 140
|||||
DB 266 AGCTCTATATTAACTTCGATGAGAGA 238

RESULT 2
BI388689 365 bp mRNA linear EST 02-AUG-2001
LOCUS 949044B01.xl 949 - Juvenile leaf and shoot cDNA from Steve Moose
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION BI388689
VERSION BI388689.1 GI:15079375
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 365)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
JOURNAL Contact: Walbot V
COMMENT Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949044 row: B column: 01.
Location/Qualifiers
1..365
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/note="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and

greening leaves 4-5 at 13 days after sowing."

BASE COUNT 108 a 81 c 70 g 106 t

ORIGIN

Query Match 7.2%; Score 29; DB 13; Length 365;
Best Local Similarity 100.0%; Pred. No. 6.7e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTCTATATTAACTTCGATGAGAGA 140
|||||
DB 264 AGCTCTATATTAACTTCGATGAGAGA 236

RESULT 3
BI396083 400 bp mRNA linear EST 08-AUG-2001
LOCUS 949044B01.yl 949 - Juvenile leaf and shoot cDNA from Steve Moose
DEFINITION Zea mays cDNA, mRNA sequence.
ACCESSION BI396083
VERSION BI396083.1 GI:15100292
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Zea.
1 (bases 1 to 400)
Walbot.V.
Maize ESTs from various cDNA libraries sequenced at Stanford
University
Unpublished (1999)
JOURNAL Contact: Walbot V
COMMENT Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 949044 row: B column: 01.
Location/Qualifiers
1..400
/organism="Zea mays"
/cultivar="W64A"
/db_xref="taxon:4577"
/clone_lib="949 - Juvenile leaf and shoot cDNA from Steve
Moose"
/tissue_type="Immature leaf primordium and vegetative
meristem"
/dev_stage="4 stages from 3-13 days after imbibing"
/lab_host="E. coli XL0LR"
/note="Organ: juvenile vegetative shoots; Vector:
PAD-GAL4-2.1; Site_1: EcoRI; Site_2: XhoI; Equal amounts
of total RNA by weight from 4 tissue sources (see below)
were pooled, polyA+ RNA isolated, and cDNA synthesized for
EcoRI (5') and XhoI (3') directional cloning into lambda
Hybridap vector from Stratagene. Tissue Sources: 1. Whole
shoots 3 days after sowing/imbibing in wet soil. 2. Basal
1.5 cm shoots 6 days after sowing - includes yellow
portions of developing leaves 1-5, primordia from 6-8, and
the vegetative apex. 3. Non-green portions of developing
leaves 4-5 and the vegetative apex, including adult leaf
primordia, 9 days after sowing. 4. Partially expanded and
greening leaves 4-5 at 13 days after sowing."

BASE COUNT 122 a 77 c 81 g 120 t

ORIGIN

Query Match 7.2%; Score 29; DB 13; Length 400;
Best Local Similarity 100.0%; Pred. No. 6.8e-05;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTCTATATTAACTTCGATGAGAGA 140
|||||
DB 55 AGCTCTATATTAACTTCGATGAGAGA 83

RESULT 4
 AM091358/c
 LOCUS AM091358 479 bp mRNA linear EST 18-OCT-1999
 DEFINITION 614094A10.xl 614 - root cDNA library from Walbot Lab Zea mays cDNA,
 mRNA sequence.
 ACCESSION AM091358
 VERSION AM091358.1 GI:6056953
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 479)
 REFERENCE Walbot,V.
 AUTHORS Malze ESTs from various cDNA libraries sequenced at Stanford
 TITLE University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 614094 row: A column: 10.
 FEATURES
 source Location/Qualifiers
 1..479
 /organism="Zea mays"
 /cultivar="W23"
 /db_xref="taxon:4577"
 /clone_lib="614 - root cDNA library from Walbot Lab"
 /tissue_type="root"
 /dev_stage="3-4 days old"
 /lab_host="XLOLR"
 /note="Organ: root; Vector: pBluescriptII SK+; Site: 1;
 EcoRI; Site: 2; XhoI; 3-4 days old root tissue from Walbot
 Lab (LM)"
 BASE COUNT 142 a 111 c 96 g 130 t
 ORIGIN
 Query Match 7.2%; Score 29; DB 10; Length 479;
 Best Local Similarity 100.0%; Pred. No. 7e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 112 AGCTTCTATATTACCTTCGATGAGAGA 140
 ||||||||||||||||||||||||||||
 Db 299 AGCTTCTATATTACCTTCGATGAGAGA 271
 RESULT 5
 BE598350
 LOCUS BE598350 487 bp mRNA linear EST 18-AUG-2000
 DEFINITION P11.81.E10.g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
 mRNA sequence.
 ACCESSION BE598350
 VERSION BE598350.1 GI:9853423
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 487)
 REFERENCE Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
 ,L.H.
 AUTHORS An EST database from Sorghum: pathogen-induced plants
 TITLE Unpublished (2000)
 JOURNAL Contact: Cordonnier-Pratt MM
 COMMENT Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: PolYTmix
 High quality sequence start: 10
 High quality sequence stop: 477
 POLYA=yes.
 FEATURES
 source Location/Qualifiers
 1..487
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (P11)"
 /note="Organ: Anthracnose-infected leaves from
 two-week-old sorghum plants 48 hr after inoculation;
 Vector: pBluescript II from Lambda Zap II; Site: 1; XhoI;
 Site: 2; EcoRI; Two-week-old sorghum plants (Brx 623
 cultivar) were infected with pathogen (isolate FRM421 of
 Colletotrichum graminicola, which is a sorghum isolate).
 RNA was prepared from infected leaves harvested from 45
 seedlings (2 weeks old) exhibit juvenile resistant
 reaction, which is an incompatible interaction. As they
 grow older (4 weeks or older), plants resume susceptibility
 to anthracnose disease. The library was made from poly-A
 RNA in the cloning vector Lambda Zap II. Clones to be
 sequenced were prepared by mass excision. WARNING: While
 most or all ESTs are expected to derive from the host
 plant, no effort was made to eliminate ESTs deriving from
 the pathogen."
 BASE COUNT 132 a 94 c 125 g 136 t
 ORIGIN
 Query Match 7.2%; Score 29; DB 10; Length 487;
 Best Local Similarity 100.0%; Pred. No. 7.1e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 112 AGCTTCTATATTACCTTCGATGAGAGA 140
 ||||||||||||||||||||||||||||
 Db 312 AGCTTCTATATTACCTTCGATGAGAGA 340
 RESULT 6
 BG558648
 LOCUS BG558648 493 bp mRNA linear EST 10-APR-2001
 DEFINITION RH122.58_E10.g1_A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
 sequence.
 ACCESSION BG558648
 VERSION BG558648.1 GI:13587646
 KEYWORDS EST.
 SOURCE Sorghum prolinguam.
 ORGANISM Sorghum prolinguam
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 1 (bases 1 to 493)
 REFERENCE Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
 ,L.H.
 AUTHORS An EST database from Sorghum: Sorghum prolinguam rhizomes
 TITLE Unpublished (2000)
 JOURNAL Contact: Cordonnier-Pratt MM
 COMMENT Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mmpratt@uga.edu
 Seq primer: T7
 High quality sequence start: 70
 High quality sequence stop: 419

FEATURES POLYA=yes.
Location/Qualifiers

source

1..493

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RHIZ2)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT

131 a 90 c 115 g 157 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.1e-05; Score 29; DB 12; Length 493;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTTCTATATTAAAGCTTCGATGAGAGA 140

DB 156 AGCTTCTATATTAAAGCTTCGATGAGAGA 184

RESULT 7

LOCUS

AW679158 535 bp mRNA linear EST 19-JUL-2000

DEFINITION WS1_22_A09_g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,

ACCESSION

AW679158

VERSION

AW679158.1 GI:7552912

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

An EST database from Sorghum: water-stressed plants

JOURNAL

COMMENT

Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmparratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: T7

High quality sequence start: 48

High quality sequence stop: 523

POLYA-No.

Location/Qualifiers

source

1..535

/organism="Sorghum bicolor"

/db_xref="taxon:4558"

/clone_lib="Water-stressed 1 (WS1)"

/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT

137 a 91 c 131 g 176 t

ORIGIN

Query Match 7.2%; Score 29; DB 10; Length 535;

Best Local Similarity 100.0%; Pred. No. 7.2e-05;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTTCTATATTAAAGCTTCGATGAGAGA 140

|||||

DB 164 AGCTTCTATATTAAAGCTTCGATGAGAGA 192

RESULT 8

LOCUS

BG842699/C 537 bp mRNA linear EST 29-MAY-2001

DEFINITION MEST39-A09_T3 ISUM4-TN Zea mays cDNA clone MEST39-A09 3', mRNA

ACCESSION

BG842699

VERSION

BG842699.2 GI:14244761

KEYWORDS

EST.

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On May 25, 2001 this sequence version replaced gi:14209021.

Contact: Patrick S. Schnable

Schnable Laboratory

Iowa State University

G405 Agronomy, Iowa State University, Ames, IA 50011-1010, USA

Tel: 515-294-0975

Fax: 515-294-2299

Email: schnable@iastate.edu

PCR Primers

FORWARD: T7-1 (AA TAC GAC TCA CTA TAG)

BACKWARD: T3 (ATT AAC CCT CAC TAA AG)

Seq primer: primer T3 (ATT AAC CCT CAC TAA AG).

Location/Qualifiers

source

1..537

/organism="Zea mays"

/cultivar="B73"

/db_xref="taxon:4577"

/clone_lib="MEST39-A09"

/clone_lib="ISUM4-TN"

/tissue.type="Seedling and silk"

/note="Vector: pT73PAC; Site_1: EcoRI; Site_2: NotI; ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with a NotI oligo-dT primer (5' AACTGAGAGATTTCGCGCGAGCAATTTTCTTTT). The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA polymerase II second strand synthesis. After the addition of EcoRI adaptors, the ds-cDNAs were digested with NotI and size-selected. The resulting molecules were directionally cloned into the EcoRI and NotI sites of the pT73PAC vector. The library then went through one round of normalization to Cot value of 5 based on the methods of Marcelo Bento Soares (Genome Research 6: 791-806, 1996)."

Research 6: 791-806, 1996."

BASE COUNT

156 a 115 c 97 g 169 t

ORIGIN

Query Match 7.2%; Score 29; DB 12; Length 537;

Best Local Similarity 100.0%; Pred. No. 7.2e-05;

Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTTCTATATTAAAGCTTCGATGAGAGA 140

|||||

DB 353 AGCTTCTATATTAAAGCTTCGATGAGAGA 325

|||||

RESULT 9

LOCUS

AW565695 568 bp mRNA linear EST 19-JUL-2000

DEFINITION LGL_348_A05_g1_A002 Light Grown 1 (LGL) Sorghum bicolor cDNA, mRNA

sequence.

ACCESSION

AW565695

VERSION AWS56595.1 GI:7219573
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 1 (bases 1 to 568)
 Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
 An EST database from Sorghum: light-grown seedlings
 Unpublished (2000)
 CONTACT: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: T7
 High quality sequence start: 29
 High quality sequence stop: 568
 POLYA=Yes.

FEATURES
 source
 1..568
 Location/Qualifiers
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Light Grown 1 (LG1)"
 /note="Organ: 10- to 14-day-old light-grown (greenhouse)
 seedlings; Vector: lambda Zap; Site.1: XhoI; Site.2: EcoRI
 ; The library was made from poly-A RNA in the cloning
 vector lambda Zap II. Clones to be sequenced were
 prepared by mass excision."

BASE COUNT 147 a 99 c 142 g 180 t
 ORIGIN

Query Match 7.2%; Score 29; DB 10; Length 568;
 Best Local Similarity 100.0%; Pred. No. 7.3e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTTATATTAGCTTCGATGAGAGA 140
 |||||||||||||||||||||||||
 Db 225 AGCTTATATTAGCTTCGATGAGAGA 253

RESULT 10
 AI978092/c 570 bp mRNA linear EST 27-AUG-1999
 LOCUS 614022D04.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
 DEFINITION mRNA sequence.
 ACCESSION AI978092
 VERSION AI978092.1 GI:5791300
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 570)
 Walbot,V
 Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 Unpublished (1999)
 CONTACT: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 614022 row: D column: 04.

FEATURES
 source
 Location/Qualifiers
 1..570
 /organism="Zea mays"
 /cultivar="W23"
 /db_xref="taxon:4577"
 /clone_lib="614 - root cDNA library from Walbot Lab"
 /tissue_type="root"
 /dev_stage="3-4 days old"
 /lab_host="XLOLR"
 /note="Organ: root; Vector: pBluescriptII SK+, Site.1:
 EcoRI; Site.2: XhoI; 3-4 days old root tissue from Walbot
 Lab (LM)"

BASE COUNT 177 a 133 c 112 g 148 t
 ORIGIN

Query Match 7.2%; Score 29; DB 9; Length 570;
 Best Local Similarity 100.0%; Pred. No. 7.3e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTTATATTAGCTTCGATGAGAGA 140
 |||||||||||||||||||||||||
 Db 391 AGCTTATATTAGCTTCGATGAGAGA 363

RESULT 11
 AW288508/c 590 bp mRNA linear EST 16-JAN-2000
 LOCUS 618068C02.x2 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA
 DEFINITION sequence.
 ACCESSION AW288508
 VERSION AW288508.1 GI:6695400
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 590)
 Walbot,V
 Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 Unpublished (1999)
 CONTACT: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 618068 row: C column: 02.

FEATURES
 source
 Location/Qualifiers
 1..590
 /organism="Zea mays"
 /cultivar="Ohio43"
 /db_xref="taxon:4577"
 /clone_lib="618 - Inbred Tassel cDNA Library"
 /tissue_type="tassel"
 /dev_stage="tassel length from 0.1 to 2.5 cm"
 /lab_host="XLOLR"
 /note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybridzap);
 Inbred tassel library from Schmidt Lab"

BASE COUNT 182 a 138 c 112 g 157 t 1 others
 ORIGIN

Query Match 7.2%; Score 29; DB 10; Length 590;
 Best Local Similarity 100.0%; Pred. No. 7.3e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 112 AGCTTATATTAGCTTCGATGAGAGA 140
 |||||||||||||||||||||||||
 Db 330 AGCTTATATTAGCTTCGATGAGAGA 302

RESULT 12
 AM036911 592 bp mRNA linear EST 15-SEP-1999
 LOCUS 614022D04.Y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
 DEFINITION mRNA sequence.
 ACCESSION AM036911
 VERSION AM036911.1 GI:5895665
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 REFERENCE 1 (bases 1 to 592)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 UNIVERSITY
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 614022 row: D Column: 04.
 Location/Qualifiers
 1..592
 /organism="Zea mays"
 /cultivar="W23"
 /db_xref="taxon:4577"
 /clone_lib="614 - root cDNA library from Walbot Lab"
 /tissue_type="root"
 /dev_stage="3-4 days old"
 /lab_host="XLOLR"
 /note="Organ: root; Vector: pBluescriptII SK+; Site_1:
 EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
 Lab (LM)"
 BASE COUNT 159 a 120 c 142 g 171 t
 ORIGIN

Query Match 7.2%; Score 29; DB 10; Length 592;
 Best Local Similarity 100.0%; Pred. No. 7.3e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTCTATATTACCTCGATGAGAGA 140
 ||||||||||||||||||||||||||||
 Db 305 AGCTCTATATTACCTCGATGAGAGA 333

RESULT 13
 B0656076 606 bp mRNA linear EST 15-JUL-2002
 LOCUS PRC0407 Sorghum halepense rhizome cDNA library Sorghum halepense
 DEFINITION cDNA clone PRC0407, mRNA sequence.
 ACCESSION B0656076
 VERSION B0656076.1 GI:21788402
 KEYWORDS EST.
 SOURCE Sorghum halepense.
 ORGANISM Sorghum halepense.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 606)
 AUTHORS Paterson,A.H.
 TITLE Unpublished, Paterson,A.H
 JOURNAL Unpublished (2002)
 COMMENT Contact: Paterson AH
 Center for Applied Genetic Technologies
 University of Georgia
 Riverend Research Laboratory, Room 162, 110 Riverbend Road, Athens
 , GA 30602, USA
 Tel: 7065830162

Fax: 7065830160
 Email: paterson@uga.edu
 Contlig
 Seq primer: M13F/M13R.
 Location/Qualifiers
 1..606
 /organism="Sorghum halepense"
 /db_xref="taxon:4560"
 /clone_lib="PRC0407"
 /tissue_type="rhizome"
 /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
 XhoI; cDNA library was prepared from mRNA isolated from
 the apical 2-3 cm of actively growing rhizomes of Johnson
 grass (Sorghum halepense) and cloned into the EcoRI/XhoI
 sites of lambda ZAP II (Stratagene)."
 BASE COUNT 152 a 110 c 147 g 197 t
 ORIGIN

Query Match 7.2%; Score 29; DB 14; Length 606;
 Best Local Similarity 100.0%; Pred. No. 7.3e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTCTATATTACCTCGATGAGAGA 140
 ||||||||||||||||||||||||||||
 Db 196 AGCTCTATATTACCTCGATGAGAGA 224

RESULT 14
 AM745589 612 bp mRNA linear EST 19-JUL-2000
 LOCUS WS1_35.A09.g1.A002 Water-stressed 1 (WS1) sorghum bicolor cDNA,
 DEFINITION mRNA sequence.
 ACCESSION AM745589
 VERSION AM745589.1 GI:7659327
 KEYWORDS EST.
 SOURCE Sorghum.
 ORGANISM Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Sorghum.
 REFERENCE 1 (bases 1 to 612)
 AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
 ,L.H.
 TITLE An EST database from Sorghum: water-stressed plants
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Laboratory for Genomics and Bioinformatics
 The University of Georgia, Department of Plant Biology
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 583 0210
 Email: mpratt@uga.edu
 Sequences have been trimmed to exclude PolyA, vector and regions
 below Phred quality 16. The threshold for highest quality sequence
 is 20.
 Seq primer: T7
 High quality sequence start: 7
 High quality sequence stop: 612
 POLYA=Yes.
 Location/Qualifiers
 1..612
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Water-stressed 1 (WS1)"
 /note="Organ: Mix of 5-week old plants on days 7 & 8 after
 water was withheld; Vector: lambda Zap; Site_1: XhoI;
 Site_2: EcoRI; The library was made from poly-A RNA in the
 cloning vector lambda Zap II. Clones to be sequenced were
 prepared by mass excision."
 BASE COUNT 156 a 110 c 148 g 198 t
 ORIGIN

Query Match 7.2%; Score 29; DB 10; Length 612;
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 112 AGCTTCTATATTAAAGCTTCGATGAGAGA 140
 ||||||||||||||||||||||||||||||||
 Db 203 AGCTTCTATATTAAAGCTTCGATGAGAGA 231

RESULT 15
 AI691590 613 bp mRNA linear EST 02-FEB-2000
 LOCUS 606022D02.x1 606 - Ear tissue cDNA library from Schmidt lab zea

DEFINITION mays cDNA, mRNA sequence.
 ACCESSION AI691590
 VERSION AI691590.1 GI:4966734

KEYWORDS EST.
 SOURCE zea mays.
 ORGANISM zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 613)

AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
 University

JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 606022 row: D column: 02.

FEATURES
 source 1..613
 location/Qualifiers
 /organism="Zea mays"
 /cultivar="Ohio43"
 /db_xref="taxon:4577"
 /clone_id="606 - Ear tissue cDNA library from Schmidt
 lab"

/tissue_type="mixed"
 /dev_stage="ear length from 0.5 cm - 2.0 cm"
 /lab_host="XLOLR (Stratagene)"
 /note="Organ: Immature ear; Vector: PBK-CMV; Site_1: EcoRI
 ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
 lab"
 BASE COUNT 178 a 138 c 114 g 183 t
 ORIGIN

Query Match 7.2%; Score 29; DB 9; Length 613;
 Best Local Similarity 100.0%; Pred. No. 7.4e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 112 AGCTTCTATATTAAAGCTTCGATGAGAGA 140
 ||||||||||||||||||||||||||||||||
 Db 348 AGCTTCTATATTAAAGCTTCGATGAGAGA 320

Search completed: January 2, 2003, 16:44:11
 Job time : 2207 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:21:06 : Search time 3182 Seconds

(without alignments)
3676.722 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402

Sequence: 1 ccggcgccggcgcacaccgct.....ggcatacttcctcgtgttc 402

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapect 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_mus:*
- 34: em_hlg_pln:*
- 35: em_hlg_rod:*
- 36: em_hlg_man:*
- 37: em_hlg_vrt:*
- 38: em_sy:*
- 39: em_higo_hum:*
- 40: em_higo_mus:*
- 41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113.2	28.2	170226	8 AP003450	AP003450 Oryza sat
2	96.8	24.1	912	8 AY097414	AY097414 Arabidops
3	96.8	24.1	1068	8 AF275664	AF275664 Arabidops
4	96.8	24.1	1440	8 AY064136	AY064136 Arabidops
5	91.4	22.7	1312	8 AY087422	AY087422 Arabidops
6	91.4	22.7	1366	8 AF428374	AF428374 Arabidops
7	86.8	21.6	1213	8 AF305635	AF305635 Pisum sat
8	82.4	20.5	1324	8 MDPEMD1	MDPEMD1
9	43.6	10.8	31522	8 AP002050	AP002050 Arabidops
10	42.8	10.6	100975	8 AC007980	AC007980 Arabidops
11	42.4	10.5	108992	8 AP003740	AP003740 Oryza sat
12	42.4	10.5	142081	8 AP004572	AP004572 Oryza sat
13	40.8	10.1	158294	8 AC091724	AC091724 Oryza sat
14	40.4	10.0	936	6 E22116	E22116 Yeast capab
15	40.4	10.0	1978	8 YSCSYT4C	M24395 Saccharomyc
16	40.4	10.0	3536	8 SCYDL047M	Z74095 S.cerevisia
17	40.4	10.0	36687	8 SCYDL37K	Z71781 S.cerevisia
18	40	10.0	165577	8 AC095138	AC095138 Rattus no
19	39.8	9.9	196679	2 AC130153	AC130153 Rattus no
20	39.2	9.8	149131	2 AC119135	AC119135 Rattus no
21	38.2	9.5	129135	9 AL136162	AL136162 Human DNA
22	38.2	9.5	194716	2 AL136227	AL136227 Homo sapi
23	38	9.5	39520	8 SPBC26H8	AC087616 Homo sapi
24	37.8	9.4	179939	2 AC079014	AC079014 Homo sapi
25	37.8	9.4	183279	2 AC079014	AC079014 Homo sapi
26	37.8	9.4	211464	2 AC023950	AC023950 Homo sapi
27	37.4	9.3	39355	2 AC130364	AC130364 Homo sapi
28	37	9.2	625	4 AF273675	AF273675 Nyctalus
29	37	9.2	1037	11 G67201	G67201 Xg4125 KWOK
30	37	9.2	100625	9 HS1189K21	AL030997 Human DNA
31	36.6	9.1	544	6 AX186776	AX186776 Sequence
32	36.6	9.1	173285	2 AC118963	AC118963 Rattus no
33	36.6	9.1	224297	2 AC121768	AC121768 Mus muscu
34	36.4	9.1	1971	6 AX299314	AX299314 Sequence
35	36.4	9.1	2529	6 AX375232	AX375232 Sequence
36	36.4	9.1	2553	6 AX283620	AX283620 Sequence
37	36.4	9.1	58198	9 AC073524	AC073524 Homo sapi
38	36.4	9.1	111103	9 AC007129	AC007129 Homo sapi
39	36.4	9.1	113991	9 AC0096742	AC0096742 Homo sapi
40	36.4	9.1	154733	2 AC127462	AC127462 Danio rer
41	36.4	9.1	164232	2 AC067792	AC067792 Homo sapi
42	36.4	9.1	174830	9 AL805977	AL805977 Human DNA
43	36.4	9.1	188209	2 AC099649	AC099649 Homo sapi
44	36.2	9.0	123116	2 AL360234	AL360234 Homo sapi
45	36.2	9.0	167611	9 AL583859	AL583859 Human DNA

ALIGNMENTS

RESULT 1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																											
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(Gs3) genomic DNA, chromosome 1, PAC
clone:P0034C09
JOURNAL Published Only in Database (2001)
REFERENCE 2 (bases 1 to 170226)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
JOURNAL Direct Submission
Submitted (28-MAR-2001) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@ias.affrc.go.jp, URL:http://rjp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)
On Nov 12, 2001 this sequence version replaced gi:15887058.
Genes were predicted from the integrated results of the following:
GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor
(October 1998 version). The genomic sequence was searched against
NCBI Nonredundant protein database, nr
(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at
RGP. Protein homologies of the coding regions were searched against
NCBI Nonredundant protein database with BLASTP2.0. ESTs represent
the identified cDNA sequences using BLASTN 2.0 with the
corresponding DBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is
classified based on the protein name to indicate the homology level
such as same name, 'putative-' and '-like protein'. A gene without
significant homology to any protein but with EST homology (covering
almost the entire length of partial sequence) is classified as an
'unknown' protein. A gene predicted with a gene prediction program
is classified as a 'hypothetical' protein.
The orientation of the sequence is from SP6 to T7 of the PAC clone.
This sequence of P0034C09 clone has an overlap with P0519D04 (DBJ:
AP003455) at 5' end and with P0431H09 (DBJ: AP003248) at 3' end.
Detailed information on overlap and assembly quality together with
annotation of this entry is available at
http://rjp.dna.affrc.go.jp/genomeseq.html.
FEATURES
source location/qualifiers
1..170226
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0034C09"
complement(join(2898..3354,3514..3964,4062..4343,
6999..7872))
/gene="P0034C09.1"
complement(join(2898..3354,3514..3964,4062..4343,
6999..7872))
/note="contains ESTs
AU032946(S00005),C22642(S15001),D48660(S15001)
unknown protein"
/codon_start=1
/protein_id="BAB84588.1"
/db_xref="gi:18565401"
/translation="MSRLCHGGRVLLPLILVAACLDGPTGDTYDAMCAQASTIC
GGVVRYPFYLSNATRALPKYANSTFCGTEGLTICDGGGGKAVMLGNDSTYSR
IDVASLVSADADVANGTCVYSHNWTIPAPSSLLHADVGMILFEFFRCAPPAAN
APPKPSIHPLTCGENSEDAPTQSFLLPASPPLPGDLMHRCSSVYCVPGVGLPDS
ANDPAMRKDGIYASLRKGFQKSMWRDSRCLETSRGKONGKFGELCCAGLVSD
SDACSKISDSTLRAGVVGGLSAVPALGLATVEFVRKKRHKVNSSSKRLKVS
GTPRSMGDMESGVKDLQTHLEFSELELEATDSFNNRELGDGEGCTYKGTILDRG
VVAVKRLYNSTRVEQFVNEAALISLRHRHNLWFTGCTSSQRELLLYEFAVNGT
VAHLHGRRAOERALSMPRLNLINVESAAALTYLHALEPPVHRDVTNNLLADAEH
VKVADPGLSRLEPLDVTHTVSTAPQGTGYVDYEHQCYQLDKSDVYSFVGLLELS
SKPAVDITRQNEINLAGMAINRIQSOLEPVDLELGEESDPAITKKMMVAELAR
CLQNGEMRRPIKEVLGKGVQDLCEMKDGGKKKGPDPPLSPDVIHQAQMSDROTT
PNTSQ"
complement(11635..12537)
/gene="P0034C09.2"
complement(11635..12537)
/gene="P0034C09.2"
/note="contains EST C26543(C12558)
unknown protein"

/codon_start=1
/protein_id="BAB84589.1"
/db_xref="gi:18565402"
/translation="MPLILGRRRLPLILVLAASHGASNDTYTSMCLQETTCG
DVSTRPPEPSDKGTGYINGSSRRSGPGALIDDDDKPILQLNGAKRYKYNVINVS
ITNLSYLDEKVEDDSSGCPRYDHWVTAAQSGMJEFFPAGMSLDIYVFLGSGFPLP
PENIDPITCSFTIGLIGPSYVLPKQVPGNNSQCKTFEPVAVTYQMDPDGDKWRG
GYGVLLKQGFLLSVNDSRRPNPCQCEBSKRCGFSODGEGTGLCLNGVRSLRCS
SDLTGKPFYSRFALYLQRPGLI"
complement(14885..15808)
/gene="P0034C09.3"
complement(14885..15808)
/gene="P0034C09.3"
/note="contains ESTs AU100785(C50325),C26880(C50325)
unknown protein"
/codon_start=1
/protein_id="BAB84590.1"
/db_xref="gi:18565403"
/translation="MAHLPLLLSFLILGVHVASDGSPLPHTYNTSICSKYKCG
VNISYPRYSNATGEYDYTOFSGCYDLNATCSMDGSKQPFQLNGDNTYILEIY
DSRTIVLVDLALRGSGCPRYRHNITFSQADEMLQYPRDLNLFPGCNDLVPPLPD
PGLFRVDKNOINCKDSNMPDGGDSFVTSVELEAMEYELASRCROYIVPVNGSI
LNSSQSALPSGEYGVLYKGFELAKMSRRDKOCLDQSRGCAYSQNTFLGCLLA
DKVYSTDCISRSKFPDAPSFSPFPCKAN"
complement(2189..22951)
/gene="P0034C09.4"
complement(2189..22951)
/gene="P0034C09.4"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB84591.1"
/db_xref="gi:18565404"
/translation="MPPPLLLPVLLAFAADASPACRNATCGDVPAVPEWLIDN
SSALGCGYGLGRVCDNTLLILPFRHORYRLDIEGKHTILSTDADMOHSDNNCS
PHILANLITDMSWLLQASSSDNTIFPKNSKSLSSVNLTCGGLGPEHIGSS
YVPLDGTITGEAVRYCEYVAVPAVIDAHKAKMADAGLIPADSGSDYLSGFEET
YSAHSDKCGCKERTKMGCGYRHNHTSTTMDFCTCDDGCPKRSKCHTAASPLTMSLL
FGVLLASDFRIGAVDIEFLIVERLKNINPRTINFSKOLLH"
join(23815..23822,24395..24764)
/gene="P0034C09.5"
join(23815..23822,24395..24764)
/gene="P0034C09.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB84592.1"
/db_xref="gi:18565405"
/translation="MRSGRDDCGHQAARAGAROSSGGRARORSRGSROAERTAS
EBACTMSTGYTEMRAATMASASSGAAPFQOTGHEVWDESHMLHSMWNCPQAEICL
TFSPSSTARQTTLNDNAVSET"
complement(join(25181..25819,25829..26579,26587..26790,
26825..26964))
/gene="P0034C09.6"
complement(join(25181..25819,25829..26579,26587..26790,
26825..26964))
/gene="P0034C09.6"
/codon_start=1
/product="similar to Oryza sativa chromosome 1,
P0031E09.21"
/protein_id="BAB84593.1"
/db_xref="gi:18565406"
/translation="MDSGCILDFRGRRGPRITGSGAEPGPBARERRPIRARCG
REKRGPLVGDHAGMGAREYRRKGFCCAMLRSGAPGRMAAGRAARLAVDDAHR
EDGGRGLTGILARRRGNVATYRAGGRKGRPMTADGGRVHNSKNGGLRANHNS
DEGEYKRRASDSOEQVRRMMNSTAAGBERKKSPPGDEDSTKQWMSLSGQTHF
RGRMGLGAAVEVRASAGSGGROWRNGHGMRLSLRLRAFCATLEPRKLEGSTFV
SRGEOBGGTTEPRGRDDOPTAKRLARRAGGCEPDGSPSVTPRPHMLPRCSA
MDGAGSARAATTLARRRERKREKRLCGAARRMGMSVLRGKFTSNAYVKGRSRTYR
QGVVTGGRGANVPYGERAKGEGKALPMPFEREGKREDASGSWTHAAMGARVYA
GGCAERRRRGRAMAAMARLPLGILTRRSGMPRGERADAGGGVLDFFERRRE
IGKAIVASARAGERERKQRESEERERKKRQDASISLGVCSARAGRRSGREERGRAR
EKRPTRKRAI"
complement(28952..29269)
/gene="P0034C09.7"

	gene	complement(29852). .29269) /gene="P0034C09.7" /note="hypothetical protein" /codon_start=1 /protein_id="BAB84594.1" /db_xref="GI:18565407"
CDS	gene	join(32474..32478,33007..33659,33836..33861) /gene="P0034C09.8" join(32474..32478,33007..33659,33836..33861) /gene="P0034C09.8" /note="contains ESTs C26880(C50325), AU100785(C50325) unknown protein" /codon_start=1 /protein_id="BAB84595.1" /db_xref="GI:18565408"
gene	gene	/translation="MDLDGSKQPFPIOLNBNYTLIEIYDSRTVIADTDALRGSCP RVRNATFGAYPMWLOXYTNSRDNLTFFEGCLINPVIDPLVSLADKHIOCNDFSN WPDSDSFEFTSGCELVAVESIELARCCSOYIVPVNPNSSINSQSALPSGGYGQVLK MGFLNNMSSDEQCCEQKCGHCHSYQNRAFLGCGLCSQGKCTKCRNNGANSSA LNMFAYOM"
CDS	gene	join(34381..35201,36446..37107,37314..37375) /gene="P0034C09.9" join(34381..35201,36446..37107,37314..37375) /gene="P0034C09.9" /note="contains ESTs C26494(C12436), AU057321(S21346)" /codon_start=1 /product="receptor serine/threonine kinase like protein" /protein_id="BAB84596.1" /db_xref="GI:18565409"
		/translation="MLSLICLRLLPLLILVLAASHGDASDPDYTSKLCOKPTTCGNVN SISVPFPAETKRDINGSNSCYGPGLAIDDDGKPILOLNGTEKRYVNTINGSTIT NLSIALDELVDSSGCPBYDHNTYIPQISMLEFEGISVDLYEFLRSFTFAKPANNA FNVAQGSFIILLTRPVEFPDELYPCGMNSQLCEETEVEPYLKQLEMDSNGMANNN SGYQVLNRQGFOLS VNBSRRPPNCSOESEGRGYQAQEFICLPNGVRSLRGN LSDRQAVAKMLKDSKGDEEFINEVASISPTSHNVNVTLLGFCIHGSKRYLIYEYMP NSLERVAFRNNSGEHSLSLWEKLEFDVVVGJARGLEYLHRCNTRIYHFIDIKPHNILLL DOECPRKSIDSCMAKLCSNKESIISIAGARETGIVIAPEVYSROFGAISKSQVSYSG MMILEMGARPRNDANSSESHFPWMYIHLDEYCINSEIDVLYNCKTYLAVTKC
Query Match		28.2%; Score 113.2; DB 8; Length 170226;
Best Local Similarity		73.2%; Pred. No. 4.1e-22;
Matches 186; Conservative		0; Mismatches 63; Indels 5; Gaps 3;
Oy	143	GAAAGAGATGTCAAGGTCTTACACAGAACCCAGAAAAACAACGAGATTCCAGGCCCAAG 202
Dd	79830	GGAAGAAGAGGTTCAATTTCTTCCAGACACAGAGAGATAAACCAGATGCGAGCCAG 79889
Oy	203	GACTTGGAATCCCATTATTTCTTCTTACT--TCAGAGTGACCAATTTTGAAGGCAATATT 259
Dd	79890	GAC-TGCTGTCCCATATTTCTCTCTACTTCATCAATGTATGACATTTTTCAGAGTC-ATGTT 79947
Oy	260	TAAATCTGGAAAGGGTAAAGAACCAATTCAGATGGTGTATATGTGGAGGTTGCAGATCAA 319
Dd	79948	TTTACTGTGAAGGGTTAAAGATGAGACGAGACTATTATTATGTTGAGAGTCCAGGTCATAG 80007
Oy	320	TGCCATCTGATGCATTCATCTTCTTGTTTTTAACAACAAGTACAAAACCAACGAGG 379
Dd	80008	TGCCATCTTAATTCGTATGATTTTCTGTTCTTTTAACACAGATGTACAATAGACCATTTGG 80067
Oy	380	AGGGGCAATCACTTC 393
Dd	80068	ACTGTCAATCACTCC 80081
RESULT 2		
LOCUS	AY097414	912 bp mRNA linear PLN 05-MAY-2002
DEFINITION	Aribidopsis thaliana AT3g19980/MZEL19_3 mRNA, complete cds.	
ACCESSION	AY097414	
VERSION	AY097414.1	GI:20453382

KEYWORDS FUL_CDNA.

SOURCE Arabidopsis thaliana.

ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE Authors
1 (bases 1 to 912)
Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Shin,P., Banh,J.,
Bower,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Arabidopsis ORF clones

TITLE Unpublished

JOURNAL 2 (bases 1 to 912)

REFERENCE Cheuk,R., Chen,H., Kim,C.J., Meyers,M.C., Shin,P., Banh,J.,
Bower,L., Carninci,P., Chang,E., Dale,J.M., Goldsmith,A.D.,
Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Karlin-Neumann,G.,
Kawai,J., Lam,B., Lee,J.M., Lin,J., Miranda,M., Narusaka,M.,
Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Sakurai,T.,
Satou,M., Seki,M., Southwick,A., Tang,C.C., Toriumi,M., Wu,H.C.,
Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W.,
Theologis,A. and Ecker,J.R.
Direct Submission

AUTHORS Submitted (15-Apr-2002) Salk Institute Genomic Analysis Laboratory
(SIGNAL), Plant Biology Laboratory, The Salk Institute for
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,
USA

COMMENT RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL_CDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members constructed and
sequenced the PUN1 (ORF) clones using the RAFL cDNAs: Cheuk,R.,
Chen,H., Kim,C.J., Meyers,M.C., Shin,P., Banh,J., Bower,L.,
Chan,H.M., Chang,E., Dale,J.M., Deng,J.M., Goldsmith,A.D.,
Jones,T., Karlin-Neumann,G., Lam,B., Lee,J.M., Lin,J., Miranda,M.,
Nguyen,M., Onodera,C.S., Palm,C.J., Quach,H.L., Southwick,A.,
Tang,C.C., Toriumi,M., Wu,H.C., Yamada,K., Yamamura,Y., Yu.G.,
Davis,R.W., Theologis,A., and Ecker,J.R.

Cheuk,R. (SSP/Salk) and Seki,M. (RIKEN GSC) contributed equally to
this work. Shinozaki,K. (RIKEN GSC) and Ecker,J.R. (SSP/Salk)
contributed equally to this work as PIs.

FEATURES Location/Qualifiers

source 1..912
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="3"
/clone="U21104"
/note="This clone is in puni 51
ecotype: Columbia"

CDS 1..912
/note="serine/threonine protein phosphatase, putative"
/codon_start=1
/product="ATPG19980/MZE19_3"
/protein_id="AAI19930.1"
/db_xref="GI:20453383"
/translation="WDQDQNTSKYKDGHLSDEDLCEVKEILLIESWQVNSGS
VTVCGDHGFHDLMKLFTGTGHPVDITVTFMGFVBDGVTSLSVFITLLKRLRPVH
NITLLRGNHESRQLTVGYVEDCCRGKNAMRWCTVDVDTLTSLAIDGVLAVCHV
GLSPEDVTIDQIRLIERNCEIPHEGPCDMLMSDPEDIEFWASPRAGWLFGSRVYAL
TEFNHINKLDLVCAHOLVOEGRLKYMFPDKGLTYWSAPNYCYRCGNVAISLSPNDNN
EREKFFETFEENNMGRPCRGVYPL"

BASE COUNT 250 a 181 c 221 g 260 t

ORIGIN

RESULT 3			
AF275664	1068 bp	mRNA	linear
LOCUS			PLN 02-JUL-2001
DEFINITION	Arabidopsis thaliana	serine/threonine protein phosphatase (STP)	

KEYWORDS	SOURCE	ORGANISM
Arabidopsis thaliana.		
Arabidopsis thaliana		

REFERENCE AUTHORS TITLE
1 (bases 1 to 1068) Kang, J.G., Kim, D.-H. and Park, C.-M. Isolation of cDNA clone encoding a putative serine/threonine

AUTHORS Kang, J. G., Kim, D. - H. and Park, C. - M.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-2000) Kumho Life & Environmental Science

```

1 1068 /organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
gene

```

```

/feature_start=1
/codon_start=1
/product="serine/threonine protein phosphatase"

```

VTVCGDJHGGFADLMKFLQTCGVCAPBDNYIENFGDVGDNSELEVFILLLLKARYPE
NTILLGNHSESRLTOVYGGYDCORRGYANAMRWYCDVEDVLTSLIIDGVLCVCA
GGLSPDYRTIDDIRLERNCETLPHEGGFCDLNMSDPEDIEYNAVSPGAGWTFGSRVY

ORIGIN	Query Match
24.18;	Score 96.8; DB 8; Length 1068;

OY 5 GCGGGGCTCACCAGCTTGGCCCAAGTGAAAGGCCTTAAGAATGTTCACAGAGGAAGGGGCCCTT 64
|||||
dh 789 GCGGGGGGCTAACACTTTCTAACAAGAGCGCTTTTACCTCATATGCTCCCAAGATTAT-TAGCCCCCT 947
|||||

65 GAACCTGTGTGGTCTGCACCCATAATTAATGCTCAGATGTGGCATGGAGCTTCTATATTA 124

125 AGCTTCATGAGAAAGAAAGAGATGTCAGGTCTTCACAGAGACCGAGAAAAA 184

Db 967 TCAATGAGAGGGCCCAAGGAC-TGGAGTCCCTATTTCCTATGA 1009

LOCUS	AY064136	1440 bp	mRNA	linear	PLN 05-DEC-2001
DEFINITION	Arabidopsis thaliana AT5g19980/MZE19_3 mRNA, complete cds.				
ACCESSION	AY064136				

ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Eumariotata; Magnoliophyta; Eudicotyledons; Core eudicots;
Rosales; Rosaceae

AUTHORS
 Baah, F., Chien, A., Chelva, R., Alam, C., U. Roesslein, E., Meyers, M.,
 Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J. M.,
 Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,
 ...

TITLE	Author
Arabidopsis cDNA clones	Yu, S., Shinozaki, K., Davis, R.W., Theologis, A. and Ecker, J.R.

Banh, J., Bowser, L., Carninci, P., Chang, E., Dale, J. M., Goldsmith, A. D., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lee, J. M., Lin, J., Miranda, M.,

TITLE
 JOURNAL
 Submitted (27-NOV-2001) Salk Institute Genomic Analysis Laboratory
 Direct Submission
 Davis, R.W., Theologos, A. and Ecker, J.R.

COMMENT

Riken Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN Arabidopsis Full-length cDNA' ; Salt M. Narusawa, M. Tshida, I.

The Saik, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RFL CDNAS: Shinn, P., Chen, H.,

Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wu, H.C., Yamada, K., Yamamura, Y., Yu, G., Yu, S., Theologis, A., and Ecker, J.R.

FEATURES
source
1. 1440
Location/Qualifiers
contributed equally to this work as PIS.

```

/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

```

```

/OD_A1E1 = LAXON.3702
/chromosome = "3"
/clone = "RAFL09-21-A14(R21104)"

```

CDS /note="ecotype: Columbia"
184. .1095

BASE COUNT	ORIGIN
385 a	277 c 304 g 474 t

LOCUS	1312 bp	mRNA	linear	PLN 26-JUN-2002
DEFINITION	AY087422	Arabidopsis thaliana clone 35230 mRNA, complete sequence.		
ACCESSION	AY087422			
VERSION	AY087422.1	GI:21406146		
KEYWORDS	FLI-CDNA.			
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	1 (bases 1 to 1312)			
AUTHORS	Haas,B.J., Volkovsky,N., Town,C.D., Troukhan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.			
TITLE	Full-length messenger RNA sequences greatly improve genome annotation			
JOURNAL	Genome Biol. (2002) In press			
REFERENCE	2 (bases 1 to 1312)			
AUTHORS	Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.			
TITLE	Full-length cDNA from Arabidopsis thaliana unpublished			
JOURNAL	3 (bases 1 to 1312)			
REFERENCE	Brover,V., Troukhan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.			
AUTHORS	Direct Submission			
TITLE	Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA			
JOURNAL	This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unsplined introns and spliced exons; one percent may contain premature stop codons; five percent may have			
COMMENT				

frame shifted in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the *ns* or *lar* ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Geneset carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.

FEATURES	location/Qualifiers
source	1. 1312 /organism="Arabidopsis thaliana" /db_xref="taxon:3702" /clone="35230"
CDS	179. .1090

Query Match	22.7%	Score 91.4	DB 8	Length 1312
Best Local Similarly	70.4%	Pred. No. 7.5e-16		
Matches 164	Conservative	0	Mismatches 66	Indels 3
				Gaps 3

Db	1048	TCMAATGAGAGGCCAAGGAC-AGGCTGTCCTATTTCCTCGAATTCATGATG	1099
RESULT 6	AF428374	1366 bp	mRNA linear
LOCUS	AF428374		PLN 18-OCT-2001
DEFINITION	Arabidopsis thaliana AtLg50370/Ft413_10	complete cds.	
ACCESSION	AF428374		
VERSION	AF428374.1	GI:16226939	
KEYWORDS	FLI-CDNA.		
SOURCE	Arabidopsis thaliana.		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 1366)		
AUTHORS	Chauk,R., Chen,H., Kim,C.J., Koesema,E., Meyers,M.C., Banh,J., Bowser,L., Carninci,P., Dale,J.M., Goldsmith,A.D., Hayashizaki,Y., Ishida,J., Jiang,P.X., Jones,T., Kamiya,A., Karlin-Neumann,G., Kawai,J., Lam,B., Lee,J.M., Lin,J., Liu,S.X., Miranda,M., Narusaka,M., Nguyen,M., Onodera,C.S., Palm,C.J., Pham,P.K., Quach,H.L., Sakurai,T., Satou,M., Seki,M., Southwick,A., Tang,C.C., Toitum,M., Yamada,K., Yamamura,Y., Yu,G., Yu,S., Shinozaki,K., Davis,R.W., Theologis,A. and Ecker,J.R.		
TITLE	Arabidopsis cDNA clones		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 1366)		

[illegible]

FEATURES	source
<p>AUTHORS TITLE JOURNAL MEDLINE REFERENCE AUTHORS TITLE JOURNAL</p>	<p>Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamiu,E. and Tabata,S. Structural analysis of Arabidopsis thaliana chromosome 3. II. Sequence features of the 4,231,695 bp regions covered by 90 P1, TAC and BAC clones DNA Res. 7 (3), 217-221 (2000)</p>
<p>COMMENT</p>	<p>2 (bases 1 to 31522) Kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamiu,E. and Tabata,S. Direct Submission Submitted (16-MAY-2000) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934) Address for correspondence: kaos@kazusa.or.jp For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/asd/graph.cgi?c-MZ19 Genes with similarity to proteins in the databases are described in 'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'. The software programs used to predict genes include: Grail (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Grail-1.3/), GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html), NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and SplicePredictor (Volker Brendel, Stanford University, http://gremlini.zoel.iastate.edu/cgi-bin/sp.cgi). Genes encoding tRNAs are predicted by tRNAscan-SE (Gean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SE/). This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions. The 5' clone is MP99 and the 3' clone is MAL21. Location/Qualifiers 1..31522</p>
<p>CDS</p>	<p>/organism="Arabidopsis thaliana" /strain="Columbia" /db_xref="taxon:3702" /chromosome="3" /clone="MZE19" /clone_1fb="Mtsui P1" join(1527..1865,2003..2104,2187..2330,2455..2605, 2715..2742,2990..3155,3270..3333,3437..3450,3552..3633, 4013..4116,4218..4356,4652..4770,5097..5249,5447..5533, 5673..5750,5911..6069,6192..6398,6627..6832,6920..7057, 7133..7262,7375..7521,7626..7693,7805..8476) /note="gene.id:MZE19.1 pir1 S33812" /codon_start=1 /evidence="not_experimental" /product="myosin-like protein" /protein_id="PAB03161.1" /db_xref="GI:119947711" /translation="MSQKVPYPMOSLSKLADRYRFDSPSPSDRLSNSGASVRLTNSNS VPRGSGLRNVSRTDIAAGSEDSYSGHGFEVDESLTDVDSGATMPLPDSERD WSPISAVARRKILQSWIOLPNGNMLKILSTGSEESVLSLPEKVKYISLTPVNPAN PILIDGVDLMLQSLYNEPSVLYLNTRYNDMSITAGSVLAVNFEPLKPLGNTRY IDAVRRKSNESPHYVYADTADIAIREMIDNOMITAGSVLAVNFEPLKPLGNTRY GGSGEYELITKTPNLPILAFGCKATTLANDMSRKKILTHFSGSKISAOIOTFILL EKSRYVOCAGEERSYHITFYOLCSAGFAMLAALMLGNVSFTYIDENHVEPAVDESLSTV ERHPTVEALDIYHVSKEDESVFAMLAALMLGNVSFTYIDENHVEPAVDESLSTV AKLIGAGINNELTLTLSKRNMRVRNDTIVOKLTLOAIDARDALAKSYSCFLFWLVED INSLGACINNETGRGSIILDIYGFESFDKSNSEDFCINTANERLQGHNRHLFLQED EYIODGIDMTVRVEDENONCLSLPEKPKLGLILSLDSESGTPEKDTLILANKIKOHL SNSCFRDRKGLPTVYHAYAEVYETTTGLEKRDLDHSGSIOLSSCSLILPQAS SMLQSEKPVAGPLYKAGASDQSLVATKFSQLEFQMLQGLGTTPTHTFTRICKIPNVI QSPRYEBOGLVLOLRCCGCVLEAVRSPPTMSHQKSRAGFLVLENIADREL OSVVALIHQRNIIPLPEMYOVGTYKLFRTSGIGVEDTRNTLHGILLEVSGSRGYQAR CLKELRKISIIOSFYRGEKIRREPELRHRRAATYIOSVSKTARLOYGIDA SVYIOSLRGIMVIRSGCDIGMLKSGCAKNIKEGVYKVSYSLEIORDIYAEALR EKREENDIIOORLOQIYENRMSSEYETKRSKMEITOKMRSLQSSLSIAKSLAVEDSA RNSDASVNSADDAIDWSSSNQFNSQISNGVSGLOPMSAGISLYIRGLAEFEBOAGVF GDQAKFLVEKVSQGVENLDDPDRRLRLKQMFEMWAKDVGRLREKTLILSKIGSESS</p>

CDS
SGSMEKVRKRMWRNRSTRY"
join(11384, .12120,12209, .12314,12486, .12676,12837, .13107)
/note="gene_id:MZE19.2
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB03162.1"
/db_xref="GI:11994772"
/translation="MGSLSGIORPLVAANAAYTASVADYSEFSESRSLRISVRSSESE
OAPSVGLVOEERSLWYSOVASKLADLSPVSRICVYPVNDILASPMCSLATSTV
SLASRVSYSALAEKASKRPATGASLIVYDTSIRGHLPEPSVDMSSSCSSEK
NRTVYLGLGSKOKHLKYADMTSGHYITFTPLPNEIMSTOVGGKAKNTESL
VNHLDWDEQKKNLVHTFSNTEGWLTYGALIEKQKODSLMGVRGCIVDSAPVA
AADPTVMSGSAAFLKKSVAATKGSASSYESNOINISOPKPAATFELLVLEKFP
AVILNLPVNRRLADVLDTLSAOPSCPOLYIYSSADRIPIVEQVESEFIVQKRAKHE
VRACNEISSPHDHSRNPETYLAELNFMNNEVFLSCCHSS"
join(13748, .13828,14115, .14261,14349, .14416,14658, .14715,
15204, .15314,15398, .15448,15554, .15697,15788, .15883,
15977, .16048,16418, .16501)
/note="gb|AAD50050.1
gene_id:MZE19.3"
/codon_start=1
/evidence-not_experimental
/product="phosphoprotein phosphatase"
/protein_id="BAB03163.1"
/db_xref="GI:11994773"
/translation="MIDLQWISKVKGQSHLSEDELQICEYVKEILLIESNVQVNSP
VTVCGDIGHQFHDMLKLFQGTGHVPTNYIEMGDEVDGYNSLEVEFTLLLKARYPA
NITLRGNHESROLTOYVGFDEQCKRYGNANAMVYCDVADYLLISLITDGYLCAH
GCLSPDVTIQIRLIERNCELPHEGPPCDLMSPEIEITWNASPRACAGLRSRVT
TEFNINKLDVCRALHOLVQSGKTFMFDKGLVTVWASPNYCYRCGNVASILSFNDNM
BREVKEFTETENNQMGRPTGVPEFL"
join(17336, .18319,18474, .18842)
/note="gene_id:MZE19.4
unknown protein"
/codon_start=1
/evidence-not_experimental
/protein_id="BAB03164.1"
/db_xref="GI:11994774"
/translation="MLIKRPNPCWRSDHHPKESNRDKMGSSSPSPDVALESARP
FLRGELKIDPKLPKSLAVLKSGVAGECWMHGSGLDHLIDITYIKIKMKAPESVCLC
GLRHSAYNSYVNLAIPTPSGTGRDYVRHVEBAESLHLHCYVPROTLIHDELFLKY
SDCELVETHDCSEVSLNNAKERGLFVGDENRKKINALVPPNGVAVKIKTGEELVNS
RRVVGFLMTMADEFDQLFGODELFCNGRLEFENLVNTALPMPGKRGKLMWNS
SRMGALYSILVREEIIMEERARAFSGFEVRAKEDDEDIDVLPVSEFCYKVLDAKE
QIEAREMYEAVYSDTSKEGYLREAREELGLCIENPVPVGHVILSOVYKGRKFE
AREBAEKGLLLLMGSPWDMKRMSEMGIAVAVYILMKSODSPDVSMGLINGLVR"
CDS
join(19425, .19551,20049, .20167,20213, .20317,20601, .20651,
20764, .20829,21160, .21261,21395, .21463,21562, .21645,
21752, .21847,21934, .21987)
/note="gene_id:MZE19.5"
/codon_start=1
/evidence-not_experimental
/product="membrane import protein-like"
/protein_id="BAB03165.1"
/db_xref="GI:11994775"
/translation="MADLPLPLTAQVADAKTKVDEKVDYSLSPVPEELHREALMS
LKSDFNEGLREDFTALNOKFSLSHRYVACISYCKTISVMAGPLEVPAOSPEETIKI
PTAHYRGANYDPKRTINEEHSOMAEFDMGSDYRAQLQGLGALIGATYOSVYN
HLSIGELTFMAGVPRKSGTGYAARTETIDKVAAGSVAGTAGVAMNYOKISDKSLAT
DEMYNYSRDAYVGYDYMRLQARVRGKIDSNVASALLERLSMGINFLLSAELDH
KKDKYKFGFLVGC"
join(23092, .23197,23303, .23979,24508, .24579,24654, .24731,
24830, .25808,26191, .26321,26566, .26787,26911, .27474,
27600, .27809,27976, .28080)
/note="gene_id:MZE19.6"
/codon_start=1
/evidence-not_experimental
/product="transcription factor-like protein"
/protein_id="BAB03166.1"
/db_xref="GI:11994776"
/translation="MDSAITSSGSDSDDEVPEOPVWPQFTFRMDPTWLSRRPLPTVD"

SHAAREHTNOAPNGCASSDTSRPGVSKREFTGNGTWNRSISSGSGADYVRLSSBOALK
RTLPPSFNSPPLPARSGTNNISNAGSVGVDYGRPRISQOALKRTLPPSPPLPSR
SGTNINRAGSGSRFGADYSPAVASVGNKSPFGDHYSGAHEIGTIOGVNGLRILPSS
LTHGTSASVILHAGSSDPMHRRFGGDEBNDRRIYQAOLVLYNPMESDLRPGT
LSVPLMRQKTAIHLAMFQKETSNNRCGILLADDOGLKTYSTALLIKOKYSOLKS
ESSCKQTEALVLADDESNAKHESGHVPELKVSNSSTVLSACGNDENSDM
EKADDEANSTRAFOKRPAACTLIYCPASVYQAKARELDEKYESKLSLVYHBS
NRTADPNELAEYDVVTYAIVTNEAPKFLVDEDEDEKNTDRYGLASGFSNNKRR
VVGAKSKSRKRGKSTNDTSEEDCGPLGVMFRVYLDEAOTIKNYRTOMARSCCT
LVARRMCTSGTPIONTDILSYFRFLRPDPVAVYSFSTIKVPISRNSCGYKRL
QAVIRAILRRTKTLTDGKRIYNLPKRVNLSQDPSVARSPYKTLLEDSRQFKA
YADGTLSONANTLILLRLQKDHQVKRYNSDPVSKVSAVRAARLPREARSRL
INRLSSAICIECNEPPEKPVYLCGHIFCYEVLLEYTGDEYTCVPRCKQOLAD
VVFESSILRNCTSDSDGSSSDHDLRSYOKRDFCSKIKAVLILQISLOPDSFN
SAOQOMPSRPPYDDDVITVEEMRLHSSSPSGAVKTIIFSQWTCMLDVELRIIE
SGIEFRRLQIDMRARIBARAVAKERSKPDYVWMLSKAGNLGNVAAACVILLDM
MNPTEOAGDARIRICGOTRVTYTRITIKOTVEDRLIKLOBEKRTWVASAFGEHSG
SSARLYVDLKITLPMV"
BASE COUNT 8810 a 5526 c 6419 g 10767 t
ORIGIN
Query Match 10.8% Score 43.6; DB 8; Length 31522;
Best Local Similarity 76.7%; Pred. No. 0.092;
Matches 66; Conservative 0; Mismatches 19; Indels 1; Gaps 1;
QY 143 GAAGAGATGTCAAGTCTTCACAGAGACCGAGAAAACAACAGATTCCAGGGCCCAAG 202
DB 16417 GGAAGGGAAGTGAAGTCTTCACAGAGACGACAGAAACATCAATGAGAGGCCCAAG 16476
QY 203 GCCTTGCAATCCCAATTTCTCTTGA 228
DB 16477 GAC-TGGAGTCCGATTTCCTATGA 16501
RESULT 10
AC007980/c
LOCUS 100975 bp DNA linear PLN 28-FEB-2001
DEFINITION Arabidopsis thaliana chromosome I BAC F1413 genomic sequence,
complete sequence.
ACCESSION AC007980
VERSION AC007980.1 GI:5391457
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 100975)
AUTHORS Federpietel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.S.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100975)
AUTHORS Federpietel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Araujo,R., Huizar,L., Rowley,D., Buehler,E., Dunn,P.,
Gonzalez,A., Kremenetskaia,I., Kim,C., Lenz,C., Li,J., Liu,S.,
Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.,
Walker,M., Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (07-JUL-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
REFERENCE 3 (bases 1 to 100975)
AUTHORS Federpietel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Brooks,S., Buehler,E., Chao,Q., Dunn,P.,
Gonzalez,A., Khan,S., Kremenetskaia,I., Kim,C., Lenz,C., Li,J.,
Liu,S., Luros,S., Schwartz,J., Shinn,P., Toriumi,M., Vysotskaia,V.,
Yu,G., Ecker,J., Theologis,A. and Davis,R.W.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,

REFERENCE	USA
AUTHORS	4 (bases 1 to 100975) Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F., Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A. and Davis,R.W.
TITLE	Direct Submission
JOURNAL	Submitted (28-FEB-2001) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA
COMMENT	e-mail for correspondence: arabsequence.stanford.edu Genes with similarity to proteins in the databases are described as putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The software programs used to predict genes include: Gtali (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/section/index.html), GENSCAN (Chris Burge, http://genomic.stanford.edu/ chris/GENSCANW.html), Fexa (V.Solovayev & A.Salamov, Sanger Centre, http://genomic.sanger.ac.uk/) and NetPlantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).
FEATURES	Location/Qualifiers
source	1..100975
gene	/organism="Arabidopsis thaliana"
CDS	/db_xref="taxon:3702"
gene	/chromosome="1"
CDS	/clone="F1413"
gene	/complement(<1..4027)
CDS	/gene="F1413.1"
gene	complement(join(<1..567,690..908,1119..1249,1596..2526, 2612..2832,2934..3087,3332..3876,3937..4027))
gene	/gene="F1413.1"
note	"Similar to transcription factors; Similar to transcription factors. This BAC contains only the 5-prime portion of the gene"
gene	/codon_start=1
CDS	/protein_id="AAD50036.1"
gene	/db_xref="GI:5734771"
CDS	/translation="MDSALDSSDVEIOETRTPOHPRIAEVVDLGASHRRDLSTLRPHLSGSSGANGHTKTGLTNDLRNFESEKPLRAEHHTHPGNGSIVSRIPN ISVGYDEKESQQAFTKRPFPESRPPEPPIDGNSGNSASHPGAGHDLGMYR NGRIILPVSAGHTSASHPHNGSLDPMHRNGIPGEINSENDERLIYQALQIMCEA LLEFLDRCGCTGPTIVAVSYFALHOCYDHLKSIYFLHNSPGTESIYQVGRCTC WSSFGSPYKSGIQVYITIRQYLSMSCVLTAVYSQKITALAMFQKENSJHCNGGILADGGLKTVSTALILKQMEAKLSKNSQNEALDLDADDESENAEFKPSKASN GSGVNDGSGIKKAKGEASTSTRKFNKRPAACTLIVCPASVVRQMARREIDKVTDEAK KLVLILYHGNGRTPDIELAKYDVTYTAIVSNEVYKOPLVDDDEDEKNSKYGIALSGFSINKRKKNVYGTTRKSKKKKNNAGSDSPDGLTAKGMPFVYVDEACTIKH PROVARACGLRAKRRMCLSGPTPIQNTIDLYSFPFLKTPPAVYKRSFHQIKPIS RNLQGLKQLQAVLRAIMLRKTKGTLIDGPIILNPPKTIINLSQVDSVEERSFYVKL ESDSRQSKFAYAAAGTLNRYANILMLLRQACHPOLKYRNSDSVGESEAEVK KLPEKEDVLSLSTRLESPPICVCYCHDEPDPVYVILGCHIFCYQCVSDYITDEPTCAP RCREOLAHNVYFSKSTLRSCVADLGCSSSEDSNHPKSYFONGEFSSTKIKAVILIDIOSLNSOGNSHTVONGMASSSQOPNDODDDDDVTYKETSLSKSTFSGNGSPITITIFS QWTGMIDVLTSLTENSIPRRRIDGMSLIARAVAKEP"
gene	complement(6533..8700)
CDS	/gene="F1413.2"
gene	complement(join(6533..6586,6685..6780,6872..6955, 7036..7104,7261..7362,7550..7615,7697..7747,8020..8149, 8208..8289,8574..8700))
gene	/gene="F1413.2"
note	"Hypothetical protein"
gene	/codon_start=1
CDS	/protein_id="AAD50049.2"
gene	/db_xref="GI:13162531"
CDS	/translation="MEGFSPPIINTAQVDATKIDKEKDYSNLPCPVLYEELREATMA LKPELFEGRFLDYKSLNOKFELSHSICIDASNSITGMPPEYANPPSSPILITIPAN YDGAAGFIDPKLDEEDKSGHIVIDYKGSYRTQIQLGNSYVANYIJOHYVPHLSL GGBAWFGQOLMSGVYAAKYTEIDKVASQSLASTVAVANNIVHAKVSEKISFEDTLY NLSRDVYTSVGYDLITTRQSLRNGKVDNSGVAAYLEEOLIGLRLLSAEVDHVRKD YKPGCVNAF"
gene	9178..9869
CDS	/gene="F1413.3"
gene	join(9178..9315,9403..9450,9539..9655,9732..9869)
gene	/gene="F1413.3"
note	"Similar to fructokinase; Similar to fructokinase, but smaller and lacking some of the sequence of fructokinase"
gene	/codon_start=1
CDS	/protein_id="AAD50037.1"
gene	/db_xref="GI:5734772"
CDS	/translation="MKATEAKEGALLSYDNLREPLMPSPPEARFTQIMSWKADI IKLLVLTGKGCRRYTKDFHGSVEFPHVADVTGAGSFGVALNQIVDDSVLEE EERLRVYLRANCGAITTKGCAIFALPTDCALSFILKROYSO"
gene	complement(10670..14375)
CDS	/gene="F1413.4"
gene	complement(join(10670..10748,10841..10964,11094..11142, 11234..11299,11516..11881,11975..12133,12413..12704, 12911..13257,13477..13671,13773..14056,14204..14375))
gene	/gene="F1413.4"
note	"Similar to oligopeptidases"
gene	/codon_start=1
CDS	/protein_id="AAD50051.1"
gene	/db_xref="GI:5734786"
CDS	/translation="MAESRSPPAKKEVHEHMEFGDVRVNDYTLNRDSTKPNMLSY LRENNYTDVMSGTQKOFENQLEFAEIRGRKEDDISAPLRKGPYYEYERKLOGKEYIO HCRLLITDNKAEPSVYDTPMGDAPPEHYIIDEENTKAOEHQVIRGARCAPSDPKV AYAEADTKGDEITYVNVVIDSEALKPVGQOLKGLISTYLEMAGNDALYITIMDELIRPKV WLHKLGTBQSDVCLTHEKDMFSLHSHSESKRYLFVASESKTYFVSLDSKTD GLVLTLPVVDGIDSSVSHRGNHEFIQIRSTEEFNSELILACPVDTISKTVLLPHRSV KIOEQLFRDLHAFERENGLOKITVHRAEQPLEGLGGNNSVFPVSDSTE SEFSRYLAKYCKSMKTPSPSYVDMDSDSYVKKIDTVLGFADASNYATERKVAAS DGOIRMSIYRNNKLLAKIDGSDPLLLYGGSYISIDVPEKASRLSLDRGFEVFAH VRGGEMGRQWYENGLKLLKNNFTPFICAEPLTELKYSCKKCKMEBSRAGLIMG AVNMRPDEKVIYAGVPYVDVLTMLDPTIPUTISEMBMDPKREERYFKYKTSIP VDVNTQNPVNLVTAGLDPVRYSEPGKWKAKREMTDNNVLLFKCGLGHSK SGRFEKLOEDAEFFAFMKVLDIMIPASG"
gene	complement(15183..17961)
CDS	/gene="F1413.5"
gene	complement(join(15183..15266,15621..15692,15782..15877, 15975..16118,16224..16274,16387..16497,16857..16914, 17131..17198,17284..17430,17881..17961))
gene	/gene="F1413.5"
note	"phosphoprotein phosphatase; Location of ESTs VBRF01, gb1233786 and ATMS2101, gb1229155. Protein has greater than 65 % identity with mammalian protein phosphatases"
gene	/codon_start=1
CDS	/protein_id="AAD50050.1"
gene	/db_xref="GI:5734785"
CDS	/translation="MDLQMSKVKRGHLSDEDLCEYVEKELLIESNNQVPSV VYCGIHGOFHDLMLKLFQGYGCHVPEPTNTIFKQDPTDRGNSLEVTILLKARHPA NTILLGNHESRQLTVYGYGDCQKRYAGNANWRCTQVDFDLTSAITIDGVLVH GGSLPVRITDQIRLIERNCEIRPHEGPFCDLMSDPEDETAVASPRGAGWLGFSVT TEFNHINLIDVLRALHQLVOEGLKVYFQKGLVTVMSAPNYCYRCGNVSIISFNDM EREKVFETFEENNRKRGPRTGVPYL"
gene	complement(19749..26167)
CDS	/gene="F1413.6"
gene	complement(join(19749..20405,20494..20561,20665..20811, 20897..21026,21113..21248,21375..21580,21772..21978, 22125..22283,22491..22568,22683..22772,22907..23059, 23130..23248,23550..23688,23778..23881,24055..24132, 24346..24409,24517..24682,24985..25039,25105..25255, 25382..25525,25614..25715,25853..26167))
gene	/gene="F1413.6"
note	"Highly similar to myosin"
gene	/codon_start=1
CDS	/protein_id="AAD50052.1"
gene	/db_xref="GI:5734787"
CDS	/translation="MAHKYKASROSILKTPADYRFGSPISDHLTENTITPPNGHLKN GVNGTSSVGGMDSVNEDPTYSRSLINKGRSSITGCGDSILPEPSLNRKQSDTYNYA RKRVITFWOQLPRGNELKIMSTSGEESVYVTEBKVLRKVSLETPVNPMDILSDVD DLMQSLVNEPAVLNIEVRYNDGMILYTAGVPLVNPKEKVEPLGNGNITAYRRAS NESPHYAIAIDFAIRMIKRVLSPPVYSITDEVQSIITISGSGAGKTEPTAKIAMOYLAA

gene	30400..32418	
Query Match	10.6%; Score 42.8; DB 8; Length 100975;	
Best Local Similarity	68.6%; Pred. No. 0.17;	
Matches	59; Conservative 0; Mismatches 27; Indels 0; Gaps 0;	
Oy	63 TTGAACCTGTGCTGCACCTATATTGCTGCATGCGCATGAGCTCTATAT 122	
Db	15700 TTGTAAAGGTGTGTGCGCCCTAATTACTGCTACCGTTGTGGAAVTGTGCTTATAT 15641	
Oy	123 TAAGCTTCGATGAGAAGAAGG 148	
Db	15640 TGAGTTCAATGACAACATGTGAGG 15615	
RESULT 11		
LOCUS	AP003740	108992 bp DNA linear PLN 03-AUG-2002
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7,	
ACCESSION	BAC clone:OJ118_B03.	
VERSION	AP003740	
KEYWORDS	AP003740.2 GI:21202839	
SOURCE		
ORGANISM	Oryza sativa (japonica cultivar-group) (cultivar:Nipponbare) DNA, clone:OJ118_B03.	
REFERENCE	Oryza sativa (japonica cultivar-group)	
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	
TITLE	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	
JOURNAL	Ehrhartoideae; Oryzaceae; Oryza.	
AUTHORS	1	
TITLE	Sasaki, T., Matsunoto, T. and Yamamoto, K.	
JOURNAL	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC clone:OJ118_B03	
AUTHORS	Published Only in Database (2001)	
TITLE	2 (bases 1 to 108992)	
JOURNAL	Sasaki, T., Matsunoto, T. and Yamamoto, K.	
AUTHORS	Direct Submission	
TITLE	Submitted (13-JUN-2001) Takuji Sasaki, National Institute of	
JOURNAL	Agrobiological Sciences, Rice Genome Research Program, Kannondai	
AUTHORS	2-1-2, Tsukuba, Ibaraki 305-8602, Japan	
TITLE	(E-mail:tsasaki@nias.affrc.go.jp, URL:http://irgp.dna.affrc.go.jp/,	
JOURNAL	Tel:81-298-38-7441, Fax:81-298-38-7468)	
AUTHORS	On May 24, 2002 this sequence version replaced gi:14422458.	
TITLE	Genes were predicted from the integrated results of the following:	
JOURNAL	GENSCAN (http://CCR-081.mt.edu/GENSCAN.html), GENESH	
AUTHORS	(http://www.softberry.com/), Genemark.hmm	
TITLE	(http://opal.biology.gatech.edu/Genemark/), Glimmer	
JOURNAL	(http://www.tigr.org/tdb/glimmer/glimt_form.html), RiceHMM	
AUTHORS	(http://irgp.dna.affrc.go.jp/RiceHMM/), SplicePredictor	
TITLE	(http://dbioinformatics.iastate.edu/cgi-bin/sp.cgi), BLASTN and	
JOURNAL	BLASTX. The genomic sequence was searched against NCBI Nonredundant	
AUTHORS	protein database, nr	
TITLE	(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at	
JOURNAL	RGP. Protein homologies of the coding regions were searched against	
AUTHORS	NCBI Nonredundant Protein database with BLASTP. ESTs represent the	
TITLE	identified cDNA sequences using BLASTN with the corresponding DBJ	
JOURNAL	accession no. and RGP clone ID.	

A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from -21M13 to M13rev of the BAC clone. This sequence of OJ118_B03 clone has an overlap with P0592C06 (DBJ:AP005197) clone at the 5' end and with P0668C05 (DBJ:AP004572) at the 3' end. The sequence was generated by combining Monsanto and RGP-Japan sequencing data. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://irgp.dna.affrc.go.jp/Genomeseq.html>.

FEATURES

source	Location/Qualifiers
LTR	1..108992
gene	/organism="Oryza sativa (japonica cultivar-group)"
misc-feature	/cultivar="Nipponbare"
	/db_xref="taxon:39947"
	/chromosome="7"
	/clone="OJ118_B03"
	complement(3276..3488)
	/note="3' LTR"
	complement(3946..8619)
	/gene="OJ118_B03.1"
	complement(3946..8619)
	/gene="OJ118_B03.1"
	/note="Probably inactive due to frameshift(s) in CDS
	probably inactive due to stop codon(s) in CDS
	pseudogene, copia polyprotein"
	/pseudo
	join(9485..9573,9643..9940,10027..10092,10185..10196)
	/gene="OJ118_B03.2"
	join(9485..9573,9643..9940,10027..10092,10185..10196)
	/gene="OJ118_B03.2"
	/note="hypothetical protein
	similar to Oryza sativa chromosome1, BAA90382"
	/codon_start=1
	/protein_id="BAC06913.1"
	/db_xref="GI:22093617"
	/translation="MRAARRMLVVTAGSTELVSAAGSTIEARRSLCAGGVWGMRA
	GEASVYRREATASTPIPVAGSGVPCVPAAGDADVASASLRAAAVAGECRRI
	EDPSAGSRANRMRKADGTILRPSVKGDGDSMRIMADSPGIDHVSQRLV"
	complement(10633..10839)
	/note="5' LTR"
	join(11450..11597,12715..12785,12991..13110,13298..13416,
	13632..13713,13847..13852)
	/gene="OJ118_B03.3"
	join(11450..11597,12715..12785,12991..13110,13298..13416,
	13632..13713,13847..13852)
	/gene="OJ118_B03.3"
	/note="contains ESTs D40795(S2947), A0081605(S2947)"
	/codon_start=1
	/product="ADP-riboseylation factor 1"
	/protein_id="BAC06914.1"
	/db_xref="GI:22093618"
	/translation="MGLAFGLFSRLFAKKEMRLVNGDAAGTTILYKIKLGEIYT
	ATDPIGVNVEYKINISFTWVDGGDKIRPIRHRVFNQGLIFVDSNDRRVVE
	ARLDHMLNDELDELAVLIVFANKODLPNMAAEITDKLGLSLRQHVYIOSTCA
	TTGGIGVEGLDMLSSN1ASKA"
	join(15518..15613,15722..15796)
	/gene="OJ118_B03.4"
	join(15518..15613,15722..15796)
	/gene="OJ118_B03.4"
	join(15518..15613,15722..15796)
	/note="hypothetical protein
	predicted by Genemark.hmm etc."
	/codon_start=1
	/protein_id="BAC06915.1"
	/db_xref="GI:22093619"

[illegible]

	/protein_id="PAC06920.1" /db_xref="GI:22093674" /translation-"MSAITTELVLDAAAGASAKAAAGTGEARAPDPRIHRRPPST SSSPREEGEPDASAGGTSFTLKTQIIFORPKVOAEWVWDPAIIIMLEKROLIVR VDSHAGEEDLPISARIARAALADDITSINYODOAQOFEFLHKMLMRELTKYONHE RESKOOKIEKRATCIPLYCLPNPTPSOTGPPIEESOEKOKEIKCAERNTLR ROSGGIELYLTKSOELSHYKSRAADMPROIILGGIRREGSAAWHFATGRSGRR CYITNGRALAVRPESHHDVDADRALLILGTIDHORANPTHHRLLLSANTLEN TLKKRKPEAVTTETKVTPESGKOLSPNNPVNYSRHRSSTTGKRLIPGGOHPTPYK VQQDRKPFAHPDDGLTSTDSETDEDKQOPPARASHORDMDNPRTKRTYOSDGK stchypa" Join(45022..45585,48876..48930,51913..51950) /gene="OJ1118.B03.11" join(45022..45585,48876..48930,51913..51950)
gene	
CDS	
Query Match	10.5%; Score 42.4; DB 8; Length 108992;
Best Local Similarity	72.4%; Pred. No. 0.22;
Matches	55; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
OY	166 CAGAGACCAGGAATAAACACCATTCAGGCCAGCGCTTGCATCCATTATTCCTT 225
Dn	DB 101614 CATACAAGAGAGAATGATCACGATGCAAGGCCAACCATTTGTTCCCGATTTCCTC 101555
OY	226 TGACTTCAGTGCGACA 241
Dn	Db 101554 TGACTTCATTTGTACA 101539
RESULT 12	
LOCUS	AF004572 142081 bp DNA linear PLN 17-AUG-2002
DEFINITION	Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 7, PAC clone:P0668C05.
ACCESSION	AF004572
VERSION	AF004572.3 GI:22296450
KEYWORDS	
SOURCE	Oryza sativa (japonica cultivar-group). (cultivar:Nipponbare) DNA, clone:P0668C05. Oryza sativa (japonica cultivar-group) Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehretoidae; Oryzaceae; Oryza.
ORGANISM	
REFERENCE	1 Sasaki,T., Matsumoto,T. and Yamamoto,K. Oryza sativa nipponbare(GAX3) genomic DNA, chromosome 7, PAC clone:P0668C05 Published Only In Database (2001)
AUTHORS	2 (bases 1 to 142081)
TITLE	Sasaki,T., Matsumoto,T. and Yamamoto,K. Direct Submission Submitted (19-Dec-2001) Takuji Sasaki, National Institute of Agrobiological Sciences, Rice Genome Research Program; Kannodai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:sasakienias.afric.go.jp, URL:http://rpg.dna.afrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468) On Aug 16, 2002 this sequence version replaced gi:21623780. Genes were predicted from the integrated results of the following: GENSCAN (http://CCR-081.mit.edu/GENSCan.html), FGENESH http://www.softberry.com/ , Genemark_hmm http://opal.biology.gatech.edu/genemark/ , glimmer http://www.tigr.org/tdb/glimmer/glmr.form.html , RICEHM http://rpg.dna.afrc.go.jp/RiceHM/ , SplicePredictor http://bioinformatics.lastate.edu/cgi-bin/sp.cgi , BIASTN and BLASTX. The genomic sequence was searched against NCBI Nonredundant Protein database, nr
JOURNAL	(ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGF. Protein homologies of the coding regions were searched against NCBI Nonredundant Protein database with BLASTP. ESTs represent the identified cDNA sequences used by BIASTN with the corresponding DDBJ accession no. and RCP clone ID.
COMMENT	A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without

significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted by two or more gene prediction programs is classified as a 'hypothetical' protein according to IRGSP standard. A gene predicted by a single gene prediction program is also classified as a probable 'hypothetical' protein and is included as a miscellaneous feature of the sequence. The orientation of the sequence is from 5' to 3' of the PAC clone. This sequence of P0668C05 clone has an overlap with OJ118_B03 (DDBJ:AP003740) clone at the 5' end. Detailed information on overlap and assembly quality together with annotation of this entry is available at <http://ftp.dna.affrc.go.jp/GenomesSeq.html>.

FEATURES

source

```
1..142081
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="7"
/clone="P0668C05"
join(3987..4169,6437..6535,7385..7510,8306..8380,
8428..8715,8759..9361)
/gene="P0668C05.1"
join(3987..4169,6437..6535,7385..7510,8306..8380,
8428..8715,8759..9361)
/gene="P0668C05.1"
/note="hypothetical protein
predicted by GENESh etc."
/codon_start=1
/protein_id="BAC10217.1"
/db_xref="GI:22296451"
/translation="MSAITTELVLVDAAAGASKAAPGTGARAAPPDRHRRPRST
SSPPREGPDASAGTSTFLKTIQIFORPKVOAFWMPYDPAIILMEKQLPVR
VDSHAGEEELPRRIARALADDTDISINOYDOAGOFHLKGMKEIRYKQHPV
REKSKOOKIEKRAKCIPLYCLPNRTSOTCRPIESQEKQKQICKEHCAGEBNL
ROSACIELYKRSOELSHYKSRNADMPRTIRGGRRRSAAARWATGSGSGRR
GYIRGRLAYRPSGHHVDADRLRLILGIDPHOANPTHRLRLKRAANLERN
TLKRPKEAVTEKRVTEPGEKQLSPNPNRTVTSNRESGTTGKRRLPGGQHPPTPK
VQORMTAHPODILTSDSGTEDDKQORPPARASHORGDNTNPKRKTKQSQGK
STCYPPA"
join(12582..13145,16436..16490,19473..19510)
/gene="P0668C05.2"
join(12582..13145,16436..16490,19473..19510)
/gene="P0668C05.2"
/note="hypothetical protein
predicted by Glimmer etc."
/codon_start=1
/protein_id="BAC10218.1"
/db_xref="GI:22296452"
/translation="MMMDHNLIPLVSSLMMLVLGLIIDVISVSNKIGRLSARF
LAHDSTVSNVYNDNDPPSSQLGGVSCFDAMVTTRLGLRMQSGEAMECQCD
IPDAIVDELDRKASGEELDAFYEDRNEDFCIASELMSYMRRLGFEQGYED
CMRMHTEFDEDRGRISYLEFRMMEDADGALYFGIEGVLIISTGCGERRAVGCV"
join(21422..21505,21590..21691,21971..22119,22452..22599)
/gene="P0668C05.3"
join(21422..21505,21590..21691,21971..22119,22452..22599)
/gene="P0668C05.3"
/note="contains ESTs D23260(C2515), C98364(C2515)"
/codon_start=1
/product="putative 60S ribosomal protein L24"
/protein_id="BAC10219.1"
/db_xref="GI:22296453"
/translation="MYRLELCRFSGAKIYPGKIGRIFRADSOVFLFSNCKRYEHNH
LKPAKLTMTMYRKQHKDIAEAVKRRRTTKRPSYSIGATLEVIQKRRARPR
RDAARESALEIRIKRTKDEKAKAEVAKSKASAGKNAAPFGKPGKGGGCKR"
join(24628..24657,25452..25481,26190..26322,27451..27491)
/gene="P0668C05.4"
join(24628..24657,25452..25481,26190..26322,27451..27491)
/gene="P0668C05.4"
/note="This category is not included in IRGSP standard.
hypothetical ORF"
```

```
predicted by GENSCAN"
join(28528..28547,28852..28939,29391..29495)
/gene="P0668C05.5"
join(28528..28547,28852..28939,29391..29495)
/gene="P0668C05.5"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by GENSCAN"
complement(join(30351..30405,31104..31280,31382..31464))
/gene="P0668C05.6"
complement(join(30351..30405,31104..31280,31382..31464))
/gene="P0668C05.6"
/note="This category is not included in IRGSP standard.
hypothetical ORF
predicted by Glimmer"
complement(40109..40452)
/note="3' LTR"
complement(40452..44645)
/gene="P0668C05.7"
complement(40452..44645)
/gene="P0668C05.7"
/note="Probably inactive due to frameshift(s) in CDS
probably inactive due to stop codon(s) in CDS
pseudogene, gypsy/Ty-3 retroelement polyprotein"
/pseudo
complement(45109..45453)
/note="5' LTR"
50849..51169
/gene="P0668C05.8"
50849..51169
/gene="P0668C05.8"
/note="hypothetical protein
predicted by Glimmer etc."
/codon_start=1
/protein_id="BAC10220.1"
/db_xref="GI:22296454"
/translation="MDPAARPSAPGCRSGTGPAAIAPHPGCRPSPAQARPPPP
PGMAPPLPLGTGPAAAPQHGPGCLGCFSRSEEEKGLDQWKKRGRDRS
EKRG"
complement(join(55139..55378,60225..60296))
/gene="P0668C05.9"
complement(join(55139..55378,60225..60296))
/gene="P0668C05.9"
/note="hypothetical protein
predicted by Genemark.hmm etc."
/codon_start=1
/protein_id="BAC10221.1"
/db_xref="GI:22296455"
/translation="WEGTRHLEKDEPPRIIDVQKRVETEMATLLEEGRGVGA
WRALEPFAEAEGKGGGADAGGIAVGAVSTGRGWMGELGGWEDGAGGWDKVG
G"
complement(join(63541..63724,63791..63849))
/gene="P0668C05.10"
complement(join(63541..63724,63791..63849))
/gene="P0668C05.10"
/note="hypothetical ORF"
/codon_start=1
/protein_id="BAC10222.1"
/db_xref="GI:22296456"
/translation="MSTVPLRHNKRELEPLNRDASVGMILMSIPILATAARPDHRS
VIGSKTIVLPTLNQIAISMEGGLDIFPANSEI"
join(66254..66289,67073..67201,67242..67324,68430..68550,
68861..69007)
/gene="P0668C05.11"
join(66254..66289,67073..67201,67242..67324,68430..68550,
68861..69007)
/gene="P0668C05.11"
/note="hypothetical ORF"
/codon_start=1
/protein_id="BAC10223.1"
/db_xref="GI:22296457"
/translation="MGIPASILTSQAQINPNQPSQRRVRRRLAGATGSCGSDSDVILR
TSAVTGGYGRARTAAASGGVRPQAAAFYGGDDQVQMLMDRSKVRSSMTSSVLS"
```

EMSCVGLGCAALISNVEMLNLLVSPILSPISLAVSLRTYTGGSANIRQLTFASN
HPSNIRILFK"
gene
70724..70885
/gene="P0668C05.12"
CDS
70724..70885
/gene="P0668C05.12"
/note="hypothetical ORF"
/codon.start=1
/protein_id="BAC10224.1"
/db_xref="GI:22296458"
/translation="MADPLAGRMHPLEKRRGEGRRDWRKGESEELARRSGRRERI
Query Match 10.5%; Score 42.4; DB 8; Length 142081;
Best Local Similarity 72.4%; Pred. No. 0.23;
Matches 55; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 166 CAGAGACCGAGAAACACAGATTCACGAGCGGCTTGATGCCATATTCCTT 225
DB 69174 CATGACAGAGAGATGACAGATGACAGGCGCACGATGTTGTCCTATTCCTC 69115
QY 226 TGACTTAGGCGACA 241
DB 69114 TGACTTCATTTGACA 69099
RESULT 13
AC091724/c 158294 bp DNA linear PLN 06-FEB-2002
LOCUS Oryza sativa chromosome 10 clone OSJNBa0004E08, complete sequence.
DEFINITION AC091724
AC091724 GI:18464007
VERSION
KEYWORDS HTG.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriarthoideae; Oryzoae; Oryza.
REFERENCE 1 (bases 1 to 158294)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Oates, R., Simmons, J., Wilson, R., Minx, P. and
Kyung, K.
Rice Genomic Sequence
Unpublished
2 (bases 1 to 158294)
Wing, R.A., Frisch, D., Presting, G., Wood, T., Yu, Y., Soderlund, C.,
Kim, H.-R., Rambo, T., Henry, D. and Simmons, J.
Direct Submission
Submitted (25-MAY-2001) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
3 (bases 1 to 158294)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Oates, R., Simmons, J., Wilson, R., Minx, P. and
Kyung, K.
Direct Submission
Submitted (26-JAN-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
4 (bases 1 to 158294)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Oates, R., Simmons, J., Wilson, R., Minx, P. and
Kyung, K.
Direct Submission
Submitted (01-FEB-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
5 (bases 1 to 158294)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Oates, R., Simmons, J., Wilson, R., Minx, P. and
Kyung, K.
Direct Submission
Submitted (03-FEB-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
6 (bases 1 to 158294)
Wing, R.A., Yu, Y., Soderlund, C., Chen, M., Kim, H.-R., Rambo, T.,
Saski, C., Henry, D., Oates, R., Simmons, J., Wilson, R., Minx, P.,

Kyung, R., Thurmond, S. K. and Sun, S.
Direct Submission
Submitted (06-FEB-2002) Clemson University Genomics Institute,
Clemson University, 100 Jordan Hall, Clemson, SC 29634, USA
On Feb 1, 2002 this sequence version replaced gi:18376840.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were confirmed by more than
one plasmid subclone; and the assembly was confirmed by restriction
digest. The following regions are single stranded and below phred
30: 120374-120415, 118002-118003 and 22860-22861. Base 22860 is
tagged ambiguous.
FEATURES
Source
Location/Qualifiers
1..158294
/organism="Oryza sativa"
/db_xref="taxon:4530"
/chromosome="10"
/clone="OSJNBa0004E08"
8731..9276
/gene="OSJNBa0004E08.1"
/note="Contains similarity to cyclophilin"
8731..9276
/gene="OSJNBa0004E08.1"
/codon.start=1
/product="Putative cyclophilin"
/protein_id="AAL75728.1"
/db_xref="GI:18542886"
/translation="MAPASSKSNPRVFLDIAIGEMVGVYIELLADKVPDAENFR
RLCTGERAGSGSKSLHYKGSAPFARVVPVCGGGITDAGNSAALDAGARHPD
EGRAVYHDDGPGVSMANAGPNTSGSFITVDAPMLDGNHVAFGVYVAGMAYRAID
RGTWSGKTYKVPVITDCYL"
16649..21066
/gene="OSJNBa0004E08.2"
/note="Contains similarity to TNP2-like transposable
element"
Join(16649..16852,17861..18860,19010..21066)
/gene="OSJNBa0004E08.2"
/codon.start=1
/product="Putative TNP2-like transposable element"
/protein_id="AAL75729.1"
/db_xref="GI:18542887"
/translation="MVNAEPNTSGSOFIIVNRPMDIGRHVRHVVDDNDIVAYER
TATWRGKMYKPVYITGLGKALALIMDRQMTADRSKEPTFCHYHILRYAENRQGE
ICPCNCKNKQKEYSASRTIHFHLESQWPSYNCWTSHEQGVTEDEVEDNDITPD
FAQYVGEQNGQTEEBEYADGNDVADLQGMQDARDCESEKARKLMDHDRTS
LYPGCEGHKRLDTTLELLQMKAKNGVSKAFQDLKLVNIIIPGKRLPETTYEAKK
IYDGLIEVYEHKIHACPMDCILYGESEENLACPYCKALRYKIRRDGPGVDQLTRK
RIPAKVWVYRPPIIPRLRIFRKNGNGMGLRMHAEFEOGDMLHPADDSQWRNIDKE
KERGKDARNIRFEPKQPDNDIVYLRVYEDLKQMKKEVPYWDDEKQEPYLRALL
EVTINDPALSINSGSNKGYKACTHCMDTESTYLKHKRVYVWGRRLANLTPVR
KKRKHFEHRADHHTKPKRHSKTVFAMVQDLKGVFGSGSDHIESEDGAAMKNS
IIFWELPYWEFLDVRHAIDVHMLTKLVNIIIPGKRLPETTYEAKK
LHDEPKRGSYHVSYPASVITLSKESKEFCEISIKVSGYSTINIKLIISMKKEFLN
LKSHDCHVMTOLIPVYIRGIIIPDNVARTITXICAPMNAISOKVINDDPRLALONEV
QCLVSEFLIRPSPFNMTILCHLYKEITILPMTLHNHPEPRTNGVTKYVRRA
RPBASLAKGSGTEVEVECEFTEDLPICVPSRHEGRARGTIGARKIMLVNHL
FRAHAFVILQSHSSIVAYIEHLLALVARNIGSDMTVTHHIDTPEYARQHLMGNE
SINQOLAEFLARGSGSIAPFOGEYINQYTFYTRADMKSTNGNSAVRVAGMGDTTA
TYGAIEDIEDIWEIDYGLPKVPLFRQOWRLGGVMDISGNTVVLNKVYSDPEFL
ANDVTOYFVFKDMSRKKGKRGDEPRQVYLETPEKSSBLRIRLRITISWNGNPL
W
16649..21066
/note="Contains similarity to TNP2-like transposable
element"
22772..23155
/gene="OSJNBa0004E08.3"
/note="Unknown protein"
22772..23155
/gene="OSJNBa0004E08.3"
/codon.start=1
/product="Unknown protein"

repeat_region	gene	CDS
repeat_region	VFAGSGSPLVDNSPDSGAKKVDGHLVAIVLGVYLGAT	
33965..33982	/note="Simple Sequence Repeat (CTG)n"	
complement(34686..35039)	/gene="OSJNBa0004E08.7"	
/note="Hypothetical protein"	complement(34686..35039)	
/gene="OSJNBa0004E08.7"		
Query Match	10.1%; Score 40.8; DB 8; Length 158294;	
Best Local Similarity	71.1%; Pred. No. 0.67;	
Matches 54; Conservative	0; Mismatches 22; Indels 0; Gaps 0;	
QY 166	CAGAGCCGAGGAAACACAGATTCAGGCGCAAGGCTTGAATCCATATTCCTT 225	
DB 77928	CATGACAGAGAGAAATGACAGATGACAGGCCACAGATGTTGTTCCCTATTCCTC 77869	
QY 226	TGACTTCAGTGGACA 241	
DB 77868	TGACTTCATTTACA 77853	
RESULT 14		
LOCUS	E22116 936 bp DNA linear PAT 18-JUN-2001	
DEFINITION	yeast capable of low-temperature regulation of activation and	
ACCESSION	E22116	
VERSION	E22116.1 GI:13023841	
KEYWORDS	JP 1999042090-A/1.	
SOURCE	Saccharomyces cerevisiae.	
ORGANISM	Saccharomyces cerevisiae	
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;	
AUTHORS	Saccharomycetales; Saccharomycetaceae; Saccharomyces.	
TITLE	1 (Phases 1 to 936)	
JOURNAL	Shun,H., Yoshinobu,K., Akio,O., Yukio,M., Tetsuji,T., Yoshie,Y.,	
COMMENT	Kozo,O., and Yuto,T.	
	Yeast capable of low-temperature regulation of activation and	
	methods for preparation and utilization thereof	
	Patent: JP 1999042090-A 1 16-FEB-1999;	
	SHOMA SANGYO CO LTD, KANEKA CORP	
	OS Saccharomyces cerevisiae	
	PN JP 1999042090-A/1	
	PD 16-FEB-1999	
	PF 29-JUL-1997 JP 1997203652	
	PR	
	PI SHUN HARASHIMA, YOSHINOBU KANEKO, AKIO OGAWA, YUKIO MUKAI, PI	
	TETSUJI TOMITA,	
	PI YOSHIE YAMAMOTO, KOZO OTAKU, YUTU TAKADA	
	PC C12N15/09, A21D8/04, A21D10/02, C12N1/19//C12N9/16, (C12N1/19, PC	
	C12N1/865),	
	PC C12N15/00	
	CC Strandedness: Double;	
	CC Topology: Linear;	
	CC Key Location/Qualifiers	
	FT CDS 1..936.	
	Location/Qualifiers	
	1..936	
	/organism="Saccharomyces cerevisiae"	
	/db_xref="taxon:4932"	
BASE COUNT	261 a 189 c 233 g 253 t	
ORIGIN		
Query Match	10.0%; Score 40.4; DB 6; Length 936;	
Best Local Similarity	59.9%; Pred. No. 0.64;	
Matches 85; Conservative	0; Mismatches 56; Indels 1; Gaps 1;	
QY 4	GCGCGGCTCACAGCTGCGCCAGTGAAGGCTTAAACATGTGCCAGAGAAAGGCGCT 63	
DB 703	GCCCGGCTCACAGCTGCTTATGGAGGGGTTCAAGTACCATTTTCTCGA-AAAGGAGCT 761	
QY 64	TGAAGTGTGTGTGCTGACACCTAATATTTGCTCAGATGTGGCAATGAGCTTATAT 123	

Db 762 TGTAAACCGTGTGGTCCGCGCCAAATTACTGTTAATAGATGTGTGTAATGTCCGAGTGTAAAT 821

Qy 124 AAGCTTCGATGAGAAAGAGAA 145

Db 822 GAAGGTGATGAGATCTTGAA 843

RESULT 1.5

LOCUS	1978 bp	DNA	linear	PLN 06-DEC-1994
DEFINITION	Saccharomyces cerevisiae protein phosphatase catalytic subunit homologue (Sit4) gene, complete cds.			

ACCESSION	M24395
VERSION	M24395.1
KEYWORDS	5'UTR; gene; homologue; protein phosphatase catalytic subunit
SOURCE	Saccharomyces cerevisiae DNA.
ORGANISM	Saccharomyces cerevisiae

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 1978)	Andt, K.T., Styles, C.A. and Fink, G.R.	A suppressor of a HIS4 transcriptional defect encodes a protein with homology to the catalytic subunit of protein phosphatases	CELL 56 (4), 527-537 (1989)

FEATURES	
source	Location/Qualifiers
	1. 1978
	/organism="Saccharomyces cerevisiae"
	/db_xref="taxon:4932"
gene	458. .1393
	/gene="ST14"
CDS	458. .1393

BASE COUNT	627 a	411 c	392 g	548 t
ORIGIN				

Query Match	10.0%	Score 40.4	DB 8	Length 1278
Best Local Similarity	59.9%	Pred. No. 0.67		
Matches 85	Conservative 0	Mismatches 56	Indels 1	Gaps 1

Accession	Sequence	Position
OY	GGCGGGGGCTACACAGCTGGGGCCAGTAAAGCCTTAAAGCATGTGCAGAGAAAGGGGCGCT	63
Db	GCCGGGCTCACCAACTGGTTATGGAGGGGTCACTACCATTTTCCTGA-AAAGGAGCT	1218
OY	TGACCTGTGTGGGTGTCGACCACTTAATTATGTGCATAGTGGCAATGGAGCTTCATATT	123
Db	TGTACCGGTGTGGGTCCGCCAATTAATCTGTTATTAGATGTGGTAATGTGCAGATGTAAAT	1278
OY	AAAGCTTGATGAGGAAGGAA	145
Db	GAAAGTGGATGAGGATCTGGAA	1300

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 11:48:16 : Search time 148 Seconds
(without alignments)
6116.914 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccggcgcgcgcgtccaccagct.....gacatactctctgtctcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: N.GeneSeq_101002:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	223.4	55.6	1444	21 AAC46559	Zea mays DNA fragm
2	93.6	23.3	1388	21 AAC47854	Arabidopsis thalia
3	91.4	22.7	1310	21 AAC47706	Arabidopsis thalia
4	91.4	22.7	1311	21 AAC38705	Arabidopsis thalia
5	89.2	22.2	1083	21 AAC50898	Arabidopsis thalia
6	86.8	21.6	912	24 ABN85631	Protein phosphatas
7	40.4	10.0	936	20 AA415651	Leu8 hPTH mutein g
8	37.2	9.3	252	14 AA036833	Human prostate exp
9	37.2	9.3	2413	23 ABV25283	

C	10	36.4	9.1	1971	24 AB150444	Human nGPCR-Seq102
C	11	36.4	9.1	2529	24 AAD29669	Human G-protein co
C	12	36.4	9.1	2553	24 AAD22616	Human G-protein co
C	13	36.2	9.0	349980	22 AAH41224	Pyrococcus abyssi
	14	35.6	8.9	234	14 AA035228	Gene for hPTH mute
	15	35.6	8.9	237	14 AA036836	Leu8 hPTH (7-84) m
	16	35.6	8.9	237	14 AA035229	Gene for hPTH mute
	17	35.6	8.9	240	14 AA035230	Gene for hPTH mute
	18	35.6	8.9	243	14 AA036829	Gene for hPTH mute
	19	35.6	8.9	252	13 AA024478	Parathyroid hormon
	20	35.6	8.9	252	14 AA036831	Gene for Cys35 hPT
	21	35.6	8.9	252	14 AA036832	Me18 hPTH mutein
	22	35.6	8.9	252	14 AA036837	Leu8, 18 hPTH mute
	23	35.6	8.9	258	24 AB149692	Human parathyroid
	24	35.6	8.9	263	13 AA024479	Parathyroid hormon
	25	35.6	8.9	745	22 AA195049	Human neuroblastom
C	26	34.8	8.7	1984	23 AAD33424	Human lung tumour
	27	34.8	8.7	9345	23 AB109624	Drosophila melanog
	28	34.2	8.5	543	22 AAH71196	Human cervical can
	29	34	8.5	18183	22 AAS46280	Tumour suppressor
	30	34	8.5	18183	24* AB170112	Chemically treated
	31	34	8.5	18183	24 ABK31159	Signal transductio
C	32	33.8	8.4	183	24 ABN91518	Staphylococcus epl
	33	33.8	8.4	438	14 AA060712	Human brain Expres
C	34	33.8	8.4	12677	22 AAK72352	Human immune/haema
C	35	33.8	8.4	12680	24 AAK72351	Human immune/haema
C	36	33.4	8.3	989	20 AAX13555	Enterococcus faeca
	37	33.2	8.3	1530	16 AA086694	Candida tropicalis
	38	33.2	8.3	3508	16 AA086693	Candida tropicalis
	39	33.2	8.2	1409	22 AAI60523	Human polynucleoti
	40	32.4	8.1	51259	18 AAX83007	Partial mouse WRN
C	41	32.2	8.0	642	23 AAK42562	Genomic sequence #
	42	32.2	8.0	642	23 ABK42563	Genomic sequence #
	43	32.2	8.0	642	23 ABK42564	Genomic sequence #
	44	32.2	8.0	642	23 ABK42565	Genomic sequence #
C	45	32.2	8.0	855	23 AAS51918	Staphylococcus aur

ALIGNMENTS

RESULT 1
AAC46559
ID AAC46559 standard; DNA: 1444 BP.
XX
AC AAC46559;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 50583.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PM EP1033405-A2.
PD
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
FR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137328.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0133724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

```
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 55.6%; Score 223.4; DB 21; Length 1444;

Best Local Similarity 79.5%; Pred. No. 1.6e-63; Mismatches 314; Conservative 0; Indels 5; Gaps 4;

```
QY 5 GCCGGGCTCACCAGCTGCGCCAGTAAAGCCTTAAGCATGTGCGAGAGGGGCTT 64
DB 1005 GCCGGGCTCACCAGCTGCGCCAGTAAAGCCTTAAGCATGTGCGAGAGGGGCTT 1063
QY 65 GACTTGTGTGTGTGTCACCTTAATTATGCTCAGATGTGCAATGAGCTTATATTA 124
DB 1064 GTCACTGTGTGTGTGTCACCTTAATTATGCTCAGATGTGCAATGAGCTTATATTA 1123
QY 125 AGCTTCAGTGAAGAAAGAGATGTCAGAGCTTCAAGACGAGGAGAAACAA 184
DB 1124 AGCTTCAGTGAAGAAAGAGATGTCAGAGCTTCAAGACGAGGAGAAACAA 1182
QY 185 CCAGATTCAGAGGCGCAGGCTTGAATCCATATTTCTTGAATCGGTGCAAT 244
DB 1183 CCAGATTCAGAGGCGCAGGCTTGAATCCATATTTCTTGAATCGGTGCAAT 1241
QY 245 TTAGAAGCAATTTAATCTGGAAGGGTAAAGAACATCAGATGTTAATTTGCG 304
DB 1242 TTAGAAGCAATTTAATCTGGAAGGGTAAAGAACATCAGATGTTAATTTGCG 1299
QY 305 AGGTTTCAGATCATATGCAATCTGATCCATTTCTGTTCTTTGAACCAAGTA 364
DB 1300 AGGTTTCAGATCATATGCAATCTGATCCATTTCTGTTCTTTGAACCAAGTA 1359
QY 365 CAAAAACCAACGAGGAGGCGATTAATCTCTGTGT 399
DB 1360 CAAATAGACCATCGGAGCAGTATGTTCTGTAT 1394
```

RESULT 2
AAC47854
ID AAC47854 standard; DNA; 1388 BP.

XX AAC47854;

XX AC

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 55364.

XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

```
XX EP1033405-A2.
PN 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
XX 18-JUN-1999; 99US-0139465.
PR 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
```


KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123160.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134236.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150866.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.7%; Score 91.4; DB 21; Length 1310;
Best Local Similarity 70.4%; Pred. No. 1e-19;
Matches 164; Conservative 0; Mismatches 66; Indels 3; Gaps 3;

OY 5 GCCGGCTCACAGCTGCGCCAGTGAAGAGCTTAAGACATGTGCCAGGAGAGGGGCGCTT 64
DB 868 GTCCGGCTCACAGCTTGTACAGAGAGGCTCTAAGTACTGTTCAGAGTAA-AGGTCTT 926
OY 65 GAACCTGTGTGCTGCACCCCTAATTATTTGCTCAGATGTGGCAATGAGCTTCTATATTA 124
DB 927 GTAACCTGTGTGCTGCCCTAATTACGTCTACCGTTGGGAAATGTGCGCTTCTATATTG 986
OY 125 AGCTTCGATGAGAAAGAGAGATGTCAAGTCTTCACAGAGACCGAGGAAACAA 184
DB 987 AGTTTCAATGAGAACATGAGAAAGG-GAAGTGAAGTTCTTCACGAGAGAGAAAGACAA 1045
OY 185 CCAGATTCACAGGCGCTTGGATCCCATTTCTCTTGTGACTTCAGTGTG 237
DB 1046 TCAATGAGAGGGCCCAAGGAC-AGGTGTTCCTTATTTCCTGTGATTCATGATG 1097

RESULT 4
AAC38705
ID AAC38705 standard; DNA; 1311 BP.
XX AAC38705;
AC AAC38705;

XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana DNA fragment SEQ ID NO: 21942.
XX
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
PN
XX
PD
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 22.7%; Score 91.4; DB 21; Length 1311;
Best Local Similarity 70.4%; Pred. No. 1e-19;
Matches 164; Conservative 0; Mismatches 66; Indels 3; Gaps 3;

QY 5 GCCGGCTCACCACTGCCCCAGATGCTTAAGCAATGTGCCAGAGAGGGCCCTT 64
Db GTCCGGCTCACCACTGTTGACAGAGATGCTCAAGTCAATGTTCCAGATTA-AGTCTT 926
QY 65 GAACCTTGTCGTCGTGACCCCTAATTAATGTCGATGGCCATGTGACGTTATATTA 124
Db TTTTCTGTGTGTGTGGCCCTAATTAATGTCACCGTTGTGGAAATGTGCTTATATG 986
QY 125 AGCTTCATGAGAGAGAGAGAGATGTCAGGCTTTCACAGAGACCGAGAAAACAA 184
Db AGTTTCATGACACATGAGAAAGG-GAAGTGAAGTTCTTCACGAGAGACAGAGAACAA 1045
QY 185 CCAGATTCAGAGCCAGAGGCTTGAATCCATATTTCCCTTGTACCTTCAGGTG 237
Db TCAATGAGAGGGCCCAAGGAC-AGGTGTCTCTATTTCTGTGTATTCATGATG 1097

RESULT 5
AAC50898
ID AAC50898 standard; DNA; 1083 BP.
XX
AC AAC50898;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 66541.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145193.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148365.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.

```

PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 22.2%; Score 89.2; DB 21; Length 1083;
 Best Local Similarity 70.9%; Pred. No. 5.1e-19;
 Matches 146; Conservative 0; Mismatches 58; Indels 2; Gaps 2;

```

QY 5 GCCGGGCTCACACGCTGGCCCAAGTGAAGGCTTAAGACATGTGCCAGAGAAAGGCGCTT 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 GTCCGGCTCACACGCTGGCTGAAGAGGTCTCAAGTACATGTCCAAAGATAA-AGGCTTT 750

QY 65 GAACCTTGCTGTGCTGCACCTTAATTATTTGCTCAGATGTGCAGATGAGCTTTATATTA 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 751 GTACTGTGTGTGTGCTGCCCTTAATTACTGTACCGTTGTGGAGATGTGCTTATATTTG 810

QY 125 AGCTTCGATGAAGAAGAAAGAGATGTCAGGCTTTCACAGACCGAGAAACAA 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 811 AGTTTCATGACAAACATGGAAGG-GAAGTGAAGTTCTTCACGAGACAGAAAGAACAA 869

```

```

QY 185 CCAGATTCAGAGCCCAAGGCGTTGGA 210
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 870 TCAATATGAGAGGCCCAAGACAGAGTA 895

```

```

RESULT 6
ABN85631
ID ABN85631 standard; DNA; 912 BP.
XX
AC ABN85631;
XX
DT 13-SEP-2002 (first entry)
XX
DE Protein phosphatase 2A catalytic subunit encoding gene.
XX
KW Protein phosphatase 2a; enzyme; plant; flowering; crop; gene; ds.
XX
OS Unidentified.
XX
FH Key location/Qualifiers
FT 1..912 /*tag= a
FT CDS /product= "Protein phosphatase 2A catalytic subunit"
XX
XX KR2001106331-A.
XX
XX 29-NOV-2001.
XX
XX 19-OCT-2001; 2001KR-0064613.
XX
XX 19-OCT-2001; 2001KR-0064613.
XX
XX 19-OCT-2001; 2001KR-0064613.
XX
XX (KOKU-) KOREA KUMHO PETROCHEMICAL CO LTD.
XX
XX Kang JG, Park CM, Song PS;
XX
XX WPI: 2002-349593/38.
XX
XX P-PSDB: ABB83820.
XX
XX Nucleic acid molecule encoding catalytic subunit of protein phosphatase
XX 2A that regulates flowering time in plants -
XX
XX PS Disclosure; Fig 2; 19pp; Korean.
XX
XX CC The invention relates to a nucleic acid molecule (ABN85631) encoding the
XX CC catalytic subunit of protein phosphatase 2A (ABB83820) that regulates
XX CC flowering timing in plants is provided, thereby delaying flowering time
XX CC in plants and transferring nutrients effectively to roots, stems and
XX CC leaves of crops.
XX
XX S0 Sequence 912 BP; 257 A; 170 C; 207 G; 278 T; 0 other;

```

Query Match 21.6%; Score 86.8; DB 24; Length 912;
 Best Local Similarity 69.5%; Pred. No. 2.9e-18;
 Matches 146; Conservative 0; Mismatches 62; Indels 2; Gaps 2;

```

QY 5 GCCGGGCTCACACGCTGGCCCAAGTGAAGGCTTAAGACATGTGCCAGAGAAAGGCGCTT 64
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 692 GTCCGGCTCACACGCTGGCTGAAGAGGTCTCAAGTACATGTCCAAAGATAA-AGGCGCTT 750

QY 65 GAACCTTGCTGTGCTGCACCTTAATTATTTGCTCAGATGTGCAGATGAGCTTTATATTA 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 751 GTAACTGTATGTGTGCTGCACCTTAATTACTGTTCACGTTGTGAAATGTGCTTATTTCTT 810

QY 125 AGCTTCGATGAAGAAGAAAGAGATGTCAGGCTTTCACAGACCGAGAAACAA 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 811 AGTTTCATGAAATATATGAGA-GAGAAAGTTAAATTTTTCACCTGAACAGAGGAACAA 869

QY 185 CCAATTCAGAGGCCCAAGGCGTTGGAATCC 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 870 TCAGATGAGAGGCCCAAGGACAGGTTGCC 899

```

```

RESULT 7
AA015651 standard; DNA; 936 BP.
ID AAX15651
XX
AC AAX15651;
XX
DT 07-MAY-1999 (first entry)
XX
DE Protein phosphatase gene coding sequence of Saccharomyces cerevisiae.
XX
KM Protein phosphatase gene; growth; fermentation activity;
XX
KM dough production; yeast; ds.
XX
OS Saccharomyces cerevisiae.
XX
PN JPL1042090-A.
XX
PD 16-FEB-1999.
XX
PF 29-JUL-1997; 97JP-0203652.
XX
PR 29-JUL-1997; 97JP-0203652.
XX
PA (KANE ) KANEKA CORP.
XX
PA (SHOS ) SHOMA SANGYO CO.
XX
DR WPI, 1999-197822/17.
XX
PT New yeast of controlled activation at low temperatures - useful for
PT improving the quality of dough
XX
PS Claim 3; Page 16-17; 41pp; Japanese.
XX
CC The present sequence represents a protein phosphatase gene sequence
CC of Saccharomyces cerevisiae. The specification describes new
CC S. cerevisiae in which the growth and/or the fermentation activity
CC is controlled at least in the range of 0-20 degrees Celsius. These
CC yeast are prepared by deleting the function of at least one protein
CC phosphatase gene. The yeast is useful in the production of dough.
XX
SQ Sequence 936 BP; 261 A; 189 C; 233 G; 253 T; 0 other;

Query Match 10.0%; Score 40.4; DB 20; Length 936;
Best Local Similarity 59.9%; Pred. No. 0.0077;
Matches 85; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

QY 4 GCGCGGCTCACACAGTGGCCAGTGAAGCTTAAGACATGCGCAGAGAGGGGCGCT 63
DB 703 GCGCGGCTCACACAGTGGCTTATGAGGGGTTCAGTACCATTTCTCTGA-AAAGGACGT 761
QY 64 TGAACCTTGTGTGCTGCAACCTTAATTATGCTCAGATGTGGCAATGAGAGCTTCTATAT 123
DB 762 TGTAACTGCTGTGCTGCGGCCAAATTAAGTATAGATGCTGTAATGTGCGCAGTGTAT 821
QY 124 AAGCTTCGATGAGAGAGGAA 145
DB 822 GAAGTGTGATGAGAGATGTGAA 843

RESULT 8
AA036833 standard; DNA; 252 BP.
ID AAQ36833
XX
AC AAQ36833;
XX
DT 09-JUN-1993 (first entry)
XX
DE Leu8 hPTH mutein gene.
XX
KM Human parathyroid hormone; hPTH; antagonist; hypoparathyroidism;
XX
KM hypercalcaemia; ss.
XX
OS Synthetic.

```

```

XX
FH Key Location/Qualifiers
FT CDS 1..252
FT mutation /*tag= a
FT 19..24
FT /*tag= b
FT /*note= "TGA ATG Leu-Met -> CTG CTG Leu-Leu"

XX
PN EP528271-A.
XX
PD 24-FEB-1993.
XX
PF 05-AUG-1992; 92EP-0113322.
XX
PR 07-AUG-1991; 91JP-0198056.
XX
PR 26-JUN-1992; 92JP-0169713.
XX
PA (TAKE ) TAKEDA CHEM IND LTD.
XX
PI Fukuda T;
XX
DR WPI, 1993-060187/08.
XX
DR P-PSDB; AAR30858.
XX
PT New human parathyroid mutein(s) - useful for treating e.g.
PT osteoporosis, hypoparathyroidism, hyperparathyroidism,
PT hypercalcaemia, hypertension etc.
XX
PS Example; Page 21; 88pp; English.
XX
CC The sequence is that of a gene coding for human parathyroid
CC hormone (hPTH) where Met8 is substd. by Leu. The peptide
CC can have higher stability, enhanced activity and improved
CC absorption by tissues. It can act as a PTH antagonist and
CC can be used as a therapeutic agent for hypercalcaemia and
CC hyperparathyroidism.
XX
SQ Sequence 252 BP; 72 A; 53 C; 65 G; 62 T; 0 other;

Query Match 9.3%; Score 37.2; DB 14; Length 252;
Best Local Similarity 56.6%; Pred. No. 0.046;
Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 65 GAACCTTGCTGCTGCACACCTTAATTATGCTCAGATGTGGCAATGAGAGCTTCTATATTA 124
DB 78 GAAGTTGCAGATGTGCACCAATTTGTTGCTTAGTGCGGCTGCTCGATGTC 137
QY 125 AGCTTCGATGAGAGAGAGAGATGTCAAGTCTTCACAGAGACCGAGAGAAACAA 184
DB 138 TGGTTCGCCAAGACACCGTAAGAGACAAATGTCTTAGTTGAGAGCCATGAAAAATC 197
QY 185 CC 186
DB 198 CC 199

RESULT 9
ABV25283/C standard; CDNA; 2413 BP.
ID ABV25283
XX
AC ABV25283;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker CDNA 25274.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
WO200160860-A2.

```

PD 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US05171.
XX
XX 17-FEB-2000; 2000US-183319P.
PR 16-MAR-2000; 2000US-189862P.
PR 25-MAY-2000; 2000US-207454P.
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
DR
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
PS Claim 1; Page 4968; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (I) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
XX Sequence 2413 BP; 565 A; 651 C; 612 G; 576 T; 9 other;
SQ
Query Match 9.3%; Score 37.2; DB 23; Length 2413;
Best Local Similarity 50.6%; Pred. No. 0.14;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
QY 217 TATTCCTTGACTTCAGGTGACAAATTTAGAAAGCAATTTTAAATCTGGAAGGGTA 276
DB 2235 TATTTTCTTTTACAGTAGTTCCTTCCGCTTATTTTATTTAGCTTTTGGGTT 2176
QY 277 AAGAACATCAGATGCTGTATTTGAGGTTCAGATCATATATGCCATCTGATGCATTC 336
DB 2175 TTATACATGACATGATATTAATGTAGAGAAACCATGTCATCTTTAGATAAATC 2116
QY 337 CATTCGTGTTCTTTTGAACAAAGTACAAAACCAACGAGGGGCATTACTTC 394
DB 2115 CATTCATTTTACCTTTTAAAAAAACAAGCTGTGTGAGACAGATGACATCC 2058
RESULT 10
ABL50444/c
ID ABL50444 standard; cDNA; 1971 BP.
XX
XX ABL50444;
XX
DT 18-JUN-2002 (first entry)
XX
DE Human ngPCR-Seq1025 polynucleotide sequence SEQ ID NO:15.
XX
XX Human; G protein-coupled receptor; GPCR; ngPCR; vulnerability; antiasthmatic;
KW anti-HIV; analgesic; cytostatic; antidiabetic; anorectic; metabolic;
KW hypertensive; hypotensive; thrombolytic; antiparkinsonian; cardiac;
KW antidiabetic; neuroleptic; antimalarial; neuroprotective; cancer;
KW tranquilizer; antidepressant; antiinflammatory; antineumatic; diabetes;

KW antiarthritic; immunosuppressive; antipsoriatic; viral infection; pain;
KW human immunodeficiency virus; obesity; anorexia; hypotension; thrombosis;
KW hypertension; myocardial infarction; cardiomyopathy; atherosclerosis;
KW Parkinson's disease; schizophrenia; migraine; anxiety; manic depression;
KW dementia; Huntington's disease; thyroid disorder; inflammatory condition;
KW rheumatoid arthritis; autoimmune disorder; hormonal disorder; psoriasis;
KW renal failure; movement disorder; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200183553-A2.
XX
XX 08-NOV-2001.
XX
XX
XX 01-MAY-2001; 2001WO-US14050.
XX
XX 03-MAY-2000; 2000US-201548P.
PR 03-MAY-2000; 2000US-201549P.
PR 03-MAY-2000; 2000US-201550P.
PR 03-MAY-2000; 2000US-201551P.
PR 03-MAY-2000; 2000US-201552P.
PR 03-MAY-2000; 2000US-201632P.
PR 03-MAY-2000; 2000US-201633P.
PR 03-MAY-2000; 2000US-201978P.
XX
XX (PHAA) PHARMACIA & UPJOHN CO.
PA
PI Parodi LA, Lind P, Sejlitz T;
XX
XX WPI; 2002-041488/05.
DR P-PSDB; ABB06813.
XX
XX
PT Novel G protein-coupled receptor-x nucleic acid molecule and
PT polypeptides encoded by them, useful for treating cancers, psoriasis,
PT Alzheimer's disease, atherosclerosis, rheumatoid arthritis, obesity,
PT anorexia -
XX
XX Claim 3; Page 70-71; 193pp; English.
XX
XX ABL50430 to ABL50496 encode the human novel G protein-coupled receptor
CC (ngPCR) proteins given in ABB0679 to ABB06867. The ngPCR sequences have
CC vulnerability, antiasthmatic, anti-HIV, analgesic, cytostatic, antidiabetic,
CC anorectic, metabolic, hypertensive, hypotensive, antidiabetic, thrombolytic,
CC antiparkinsonian, cardiac, neuroleptic, antiemetic, antimalarial, neuroprotective,
CC tranquilizer, antidepressant, antiinflammatory, antineumatic, antipsoriatic
CC activities. They can be used as immune response inducers, ngPCR
CC expression or activity modulators and in gene therapy. ngPCR
CC polynucleotide sequences can be used as probes to screen databases and
CC nucleic acid libraries, for screening restriction fragment length
CC polymorphism associated with certain disorders, as well as for genetic
CC mapping. ngPCR sequences can be used in the treatment of diseases such
CC as viral infections caused by human immunodeficiency virus (HIV) 1 or
CC HIV-2, pain, cancers, diabetes, obesity, anorexia, hypotension,
CC hypertension, thrombosis, myocardial infarction, cardiomyopathies,
CC atherosclerosis, Parkinson's disease, schizophrenia, migraine, anxiety,
CC manic depression, dementia, Huntington's disease, thyroid disorders,
CC inflammatory conditions, rheumatoid arthritis, autoimmune disorders,
CC hormonal disorders, renal failure, psoriasis, movement disorders.
XX
XX Sequence 1971 BP; 652 A; 311 C; 374 G; 634 T; 0 other;
SQ
Query Match 9.1%; Score 36.4; DB 24; Length 1971;
Best Local Similarity 54.5%; Pred. No. 0.24;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 151 ATGTCAAGTCTTCACAGACGAGGAAACCAACGATTCAGGGCCAAAGGCTTGA 210
DB 246 ATACCAAGCTGTGGAAGAGTACAGAGATGACTACAGAGCCACGACGAGCTCGGA 187
QY 211 ATCCCATATTTCCCTTGACTTCAGGTGACAAATTTTGAAGGCAATTTTAAATCTGGA 270
DB 186 CAGAGACATACCTTTGATTTGAGGTAGACATTTAGCTGATGACATTTTGTAAATTTTGA 127

Yr	271	GGGCTAAAGAACAA	284
Db	126	TATGTAACACACTA	113
RESULT 11			
AAD29669/c			
ID	AAD29669	standard; cDNA; 2529 BP.	
AC			
XX	AAD29669;		
DT	17-MAY-2002	(first entry)	
DE	Human G-protein coupled receptor (GCRC-3) cDNA.		
XX			
KW	Human; G-protein coupled receptor; GCRC-3; cell proliferative disorder;		
KW	neurological; cardiovascular; gastrointestinal; autoimmune; inflammatory		
KW	metabolic; hepatic; psoriasis; cancer; epilepsy; Alzheimer's disease;		
KW	Pick's disease; Huntington's disease; Parkinson's disease; hypertension;		
KW	atherosclerosis; myocardial infarction; gastritis; cirrhosis; cytostatic		
KW	apoptosis; Crohn's disease; acquired immunodeficiency syndrome; AIDS;		
KW	anemia; asthma; rheumatoid arthritis; diabetes; obesity; drug screening		
KW	transgenic animal; allergy; gene therapy; hepatotropic; anticonvulsant;		
KW	neotropic; neuroprotective; cardiant; immunosuppressive; anorectic;		
KW	vitucide; gene; ss.		
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FH	CDS	24..1310	
FT		/*tag= a	
FT		/product= "Human GCRC-3"	
PN	WO200210387-A2.		
XX			
XX	07-FEB-2002.		
PD			
XX			
XX	25-JUL-2001; 2001MO-US23433.		
PF			
XX			
XX	27-JUL-2000; 2000US-221478P.		
PR	03-AUG-2000; 2000US-223268P.		
PR	21-AUG-2000; 2000US-227054P.		
PR	08-SEP-2000; 2000US-231121P.		
PR	13-SEP-2000; 2000US-232243P.		
PR	15-SEP-2000; 2000US-232691P.		
PR	22-SEP-2000; 2000US-235146P.		
XX			
PA	(INCYT-) INCYTE GENOMICS INC.		
PI	Thornton M, Patterson C, Lal P, Burford N, Yue H, Gandhi AR;		
PI	Elliot VS, Ramkumar J, Baughn MR, Kallick DA, Walia NK;		
PI	Hattalia ADA, Yao MG, Lu Y, Tribouley CM, Policky JL, Kearney L;		
PI	Grail RC, Warren BA, Lee EA, Ding L;		
DR	WPI: 2002-188744/24.		
DR	P-PSDB; AAEL18642.		
XX			
XX	New human G-protein coupled receptor polypeptide for diagnosis,		
PT	prevention and treatment of cell proliferative, neurological,		
PT	cardiovascular, gastrointestinal, autoimmune/inflammatory, and		
PT	metabolic disorders		
XX			
PS	Claim 5; Page 138-139; 150pp; English.		
XX			
CC	The invention relates to novel human G-protein coupled receptors		
CC	(GCRC) and their encoding polynucleotides. GCRC is useful as an		
CC	immunogen for preparing monoclonal and polyclonal antibodies. GCRC is		
CC	useful for diagnosing, treating and preventing a cell proliferative		
CC	disorder (e.g., hepatitis, psoriasis, cancer), a neurological disorder		
CC	(e.g., epilepsy, Alzheimer's disease, Pick's disease, Huntington's		
CC	disease, Parkinson's disease), a cardiovascular disorder (e.g.,		
CC	atherosclerosis, hypertension, myocardial infarction), gastrointestinal		

CC	disorder (e.g., gastritis, cirrhosis, Crohn's disease), an autoimmune/
CC	inflammatory disorder (e.g., acquired immunodeficiency syndrome (AIDS),
CC	allergy, anaemia, asthma, rheumatoid arthritis), a metabolic disorder
CC	(e.g., diabetes, obesity, osteoporosis), and viral infections. GCRC is
CC	useful in a number of drug screening techniques, and to analyse the
CC	proteome of a tissue or cell type. GCRC is useful for creating knockin
CC	humanised animals or transgenic animals to model human diseases, in
CC	somatic or germline gene therapy, to generate a transcript image of a
CC	tissue or cell type, for detecting differences in the chromosomal
CC	location due to translocation, inversion, etc., among normal, carrier
CC	or affected individuals, and as hybridization probes for mapping
CC	naturally occurring genomic sequences. GCRC is useful in Southern or
CC	northern analysis, dot blot or other membrane-based technologies, in PCR
CC	technologies, in dipstick, pin, multiformat enzyme linked immunosorbant
CC	(ELISA)-like assays, and in microarrays utilising fluids or tissues from
CC	patients to detect altered GCRC expression. The present sequence is
CC	human GCRC-3 cDNA.
XX	
SO	Sequence 2529 BP; 726 A; 516 G; 525 G; 762 T; 0 other;
	Query Match 9.1%; Score 36.4; DB 24; Length 2529;
	Best Local Similarity 54.5%; Pred. No. 0.27;
	Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps
OY	151 ATGTCAAGTCTTCACAGAGACCGAGAAACAACAGATTCCAGGGCCAGGCGCTTGCA 210
DB	809 ATACAGACTCTGGAGAGAGTACAGAAATGACTACAGACCACGACGACACTGGCA 750
OY	211 ATCCATATTTCTTCTTACTCAGGTGACATTTTGAAGGCAATTTAAATCTGGAA 270
DB	749 CAGAGACCTACCTTTGATTCGAGTAGACATTAAGCTGATGACATTTTGTAAATTTGGCA 690
OY	271 GGGTAAAGACAA 284
DB	689 TATGTAAACACTA 676
XX	
XX	RESULT 12
XX	AAD22616/c
ID	AAD22616 standard; cDNA; 2553 BP.
XX	
XX	AAD22616;
XX	
XX	26-FEB-2002 (first entry)
DE	
XX	Human G protein-coupled receptor (GPCR) nPCR-1025 cDNA.
XX	
XX	Human; G protein-coupled receptor; GPCR; gene therapy; cancer;
KW	tumour growth; thyroid disorder; myxoedema; renal failure;
KW	inflammatory condition; Crohn's disease; cell differentiation;
KW	homeostasis; rheumatoid arthritis; autoimmune disorder;
KW	movement disorder; central nervous system disorder; stroke;
KW	Huntington's disease; Tourette's Syndrome; Parkinson's disease;
KW	Alzheimer's disease; viral infection; human immunodeficiency virus;
KW	HIV; metabolic disorder; cardiovascular disease; type 2 diabetes;
KW	obesity; hypotension; hypertension; thrombosis; myocardial infarction;
KW	atherosclerosis; proliferative disease; hyperproliferative disorder;
KW	perioritis; hormonal disorder; polycystic ovarian syndrome; alopecia;
KW	sexual dysfunction; respiratory ailment; asthma; brain injury;
KW	neotrophic; neuroprotective; cytoskeletal; antineumatic; antiarthritic;
KW	cerebroprotective; anticonvulsant; neuroleptic; anorectic; cardiant;
KW	thrombolytic; antiarteriosclerotic; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	CDS 40..1329
FT	/*tag= a
FT	/product= "Human nPCR-1025 protein"
XX	W0200179292-A2.
XX	
XX	25-OCT-2001

XX 17-APR-2001; 2001WO-US12470.
PF
XX
XX 17-APR-2000; 2000US-198090P.
PR
XX (PHAA) PHARMACIA & UPJOHN CO.
PA
XX Soderberg C, Lind P;
PI
XX WPI: 2002-026013/03.
DR
XX P-PDB; AAEL3533.
DR
XX
XX Novel isolated nucleic acid molecule encoding G protein-coupled
PT receptor polypeptide, ngPCR-1025, useful for treating, e.g., cancer,
PT Crohn's disease, rheumatoid arthritis, Alzheimer's disease, stroke,
PT thrombosis and psoriasis
PS
XX Claim 3; Page 59-60; 98pp; English.
XX
XX The present invention relates to an isolated nucleic acid molecule
CC comprising a nucleotide sequence that encodes a G protein-coupled
CC receptor (GPCR) polypeptide, ngPCR-1025. ngPCR-1025 DNA may also be
CC useful in gene therapy. The invention is used to treat or prevent
CC unregulated cellular growth, such as cancer cell and tumour growth,
CC and for treating thyroid disorders (e.g., myxoedema), renal failure,
CC inflammatory conditions (e.g., Crohn's disease), diseases related to
CC cell differentiation and homeostasis, rheumatoid arthritis, autoimmune
CC disorders, movement disorders, central nervous system disorders (e.g.,
CC stroke, Huntington's disease, Tourette's Syndrome, Parkinson's disease,
CC Alzheimer's disease), infections, such as viral infections caused by
CC HIV-1 or HIV-2 (human immunodeficiency virus), metabolic and
CC cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity,
CC hypertension, thrombosis, myocardial infarction,
CC atherosclerosis), proliferative diseases and cancers, hyperproliferative
CC disorders (such as psoriasis), hormonal disorders (e.g., polycystic
CC ovarian syndrome, alopecia), and sexual dysfunction. ngPCR-1025 is useful
CC for treating respiratory ailments such as asthma and for treating acute
CC and/or traumatic brain injury. The present sequence is a human ngPCR-1025
CC DNA.
XX
XX Sequence 2553 BP; 717 A; 540 C; 532 G; 764 T; 0 other;
SQ
Query Match 9.1%; Score 36.4; DB 24; Length 2553;
Best Local Similarity 54.5%; Pred. No. 0.27;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 151 ATGCAAGTCTTCACAGAGACGAGAAACACAGATTCAGGCGCAAGGCTTGGCA 210
DB 828 ATAAACAAGCTCTGAGAGAGTACAGAAAGATGACTACAGAGCCGACGACAGCTCTGGCA 769
QY 211 ATCCCATATTTCCCTTGAAGTTCAGGTCGACATTTTGAAGGCAATTTTAAATCTGGAA 270
DB 768 CAGAGCATATCCCTTATTCGAGGTAGACATTTAGCTGATGACATTTTGTAAATTTTGA 709
QY 271 GGGGTAAAGACAA 284
DB 708 TATGTAAACACACTA 695
RESULT 13
AAH41224/C
ID AAH41224 standard; DNA: 349980 BP.
XX
XX AAH41224;
XX
XX 29-OCT-2001 (first entry)
XX
XX Pyrococcus abyssi genomic fragment #3.
XX
XX Pyrococcus abyssi
XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.
XX Pyrococcus abyssi.

FH Key Location/Qualifiers
FT misc_feature 1..49980
FT /*tag= a
FT /note= "This sequence overlaps with the 3' end of
FT AAH41223"
FT misc_feature 300001..349980
FT /*tag= b
FT /note= "This sequence overlaps with the 5' end of
FT AAH41225"
XX
XX FR2792651-A1.
XX
XX 27-OCT-2000.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX 21-APR-1999; 99FR-0005034.
XX
XX (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Hellig R;
XX
XX WPI: 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
PT
XX Claim 1; Page 347-443; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
CC abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
CC hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
CC vents. The present sequence is a fragment of the genomic sequence of P.
CC abyssi. The 5' end of this sequence overlaps with the 3' end of AAH41223
CC and the 3' end of this sequence overlaps with the 5' end of AAH41225. The
CC proteins of the present invention have various potential industrial uses,
CC since the proteins are stable at very high temperatures, some up to 110
CC degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAH75903-AAH75920 and AAG66436.
XX
XX Sequence 349980 BP; 94090 A; 78692 C; 78319 G; 98879 T; 0 other;
SQ
Query Match 9.0%; Score 36.2; DB 22; Length 349980;
Best Local Similarity 60.8%; Pred. No. 3.8;
Matches 59; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 85 CTATATTGCTCGAGATGCGCATGAGCTTCTATATTAAAGCTTCATGAGAAAGCA 144
DB 54308 CGAAGATTTGGCCAAATGGGGAATAGAGTTCTCGAGAAAGATTCACGAGGAAGTG 54249
QY 145 AAGGAGATGTCAGAGCTTCACAGACGACGAGAAA 181
DB 54248 ATGAAGAAGTTGAGCTGACGATTAATCGAGGAAAA 54212
RESULT 14
AAO35228
ID AAO35228 standard; DNA: 234 BP.
XX
XX AAO35228;
XX
XX 09-JUN-1993 (first entry)
XX
XX Gene for hPTH mutetin lacking 6 N-terminal amino acids.
XX
XX Human parathyroid hormone; hPTH; antagonist; hypoparathyroidism;
KW hypercalcaemia; ss.
XX
XX Synthetic.
OS

```
XX Key Location/Qualifiers
FH 1..234
FT CDS /*tag= a
FT /note= "codes for peptide lacking N-terminal region"
EP528271-A.
PN 24-FEB-1993.
PD 05-AUG-1992; 92EP-0113322.
PF 07-AUG-1991; 91JP-0198056.
PR 26-JUN-1992; 92JP-0169713.
XX (TAKE ) TAKEDA CHEM IND LTD.
XX Fukuda T;
PI WPI: 1993-060187/08.
DR P-PSDB; AAR30852.
DR New human parathyroid muten(s) - useful for treating e.g.
XX osteoporosis, hypoparathyroidism, hyperparathyroidism,
PT hypercalcaemia, hypertension etc.
PS Disclosure; Page 7; 88pp; English.
XX The sequence is that of a gene coding for a human parathyroid hormone
CC (hPTH) peptide lacking 6 hPTH N-terminal amino acids. This peptide
CC can have higher stability, enhanced activity and improved absorption
CC by tissues. It can act as a PTH antagonist and can be used as a
CC therapeutic agent for hypercalcaemia and hyperparathyroidism.
XX
SQ Sequence 234 BP; 71 A; 48 C; 57 G; 58 T; 0 other;
Query Match 8.9%; Score 35.6; DB 14; Length 234;
Best Local Similarity 55.7%; Pred. No. 0.15;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
OY 65 GAACCTGTGTGTCGCACCCCTAATTATGCTCAGATGTGCAATGGAGCTTCTATATTA 124
DB 60 GAAGTTGCAGAGATGTGCACAATTTGTCCTTAGTGCCCATGGCTCCTCGTATGC 119
OY 125 AGCTTCGATGAGAGAGAGAGAGATGTCAAGGCTTCACAGACCGAGAAACAA 184
DB 120 TGGTTCCTCCAAAGACCAAGTAAAGAGACATGTCTTAGTGTAGAGCCATGAAAAATC 179
OY 185 CC 186
DB 180 CC 181
RESULT 15
AAQ36836
ID AAQ36836 standard; DNA; 234 BP.
XX
AC AAQ36836;
XX
DT 09-JUN-1993 (first entry)
XX
DE Leu8 hPTH (7-84) muten gene.
XX
KW Human parathyroid hormone; hPTH; antagonist; hypoparathyroidism;
KW hypercalcaemia; ss.
XX
XX Synthetic.
XX
OS
XX
FH Key Location/Qualifiers
FT 1..234
FT CDS /*tag= a
FT /note= "encodes muten hPTH amino acids 7-84"
FT mutation 4..6
```

```
FT /*tag= b
FT /note= "ATG Met -> CTC Leu"
EP528271-A.
PN 24-FEB-1993.
PD 05-AUG-1992; 92EP-0113322.
PF 07-AUG-1991; 91JP-0198056.
PR 26-JUN-1992; 92JP-0169713.
XX (TAKE ) TAKEDA CHEM IND LTD.
XX Fukuda T;
PI WPI: 1993-060187/08.
DR P-PSDB; AAR30852.
DR New human parathyroid muten(s) - useful for treating e.g.
XX osteoporosis, hypoparathyroidism, hyperparathyroidism,
PT hypercalcaemia, hypertension etc.
PS Example; Page 21; 88pp; English.
XX The sequence is that of a gene coding for human parathyroid
CC hormone (hPTH) amino acids 7-84 where Met8 is substd. by Leu.
CC The peptide can have higher stability, enhanced activity and
CC improved absorption by tissues. It can act as a PTH antagonist
CC and can be used as a therapeutic agent for hypercalcaemia and
CC hyperparathyroidism.
XX
SQ Sequence 234 BP; 70 A; 50 C; 57 G; 57 T; 0 other;
Query Match 8.9%; Score 35.6; DB 14; Length 234;
Best Local Similarity 55.7%; Pred. No. 0.15;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
OY 65 GAACCTGTGTGTCGCACCCCTAATTATGCTCAGATGTGCAATGGAGCTTCTATATTA 124
DB 60 GAAGTTGCAGAGATGTGCACAATTTGTCCTTAGTGCCCATGGCTCCTCGTATGC 119
OY 125 AGCTTCGATGAGAGAGAGAGAGATGTCAAGGCTTCACAGACCGAGAAACAA 184
DB 120 TGGTTCCTCCAAAGACCAAGTAAAGAGACATGTCTTAGTGTAGAGCCATGAAAAATC 179
OY 185 CC 186
DB 180 CC 181
```

Search completed: January 2, 2003, 12:29:55
Job time : 304 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:23:21 : Search time 74 Seconds
(without alignments)
1666.001 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402

Sequence: 1 ccggcgccggagctaccacagct.....ggcatacttcctgtgttc 402

Scoring table: IDENTITY_NNC

Searched: Gap0 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCRTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.6	8.9	234	2	US-08-733-446-19
2	35.6	8.9	234	2	US-08-733-446-61
3	35.6	8.9	237	2	US-08-733-446-20
4	35.6	8.9	240	2	US-08-733-446-21
5	35.6	8.9	243	2	US-08-733-446-22
6	35.6	8.9	245	2	US-08-733-446-24
7	35.6	8.9	247	2	US-08-733-446-25
8	35.6	8.9	248	2	US-08-733-446-26
9	35.6	8.9	250	2	US-08-733-446-27
10	35.6	8.9	251	2	US-08-733-446-28
11	35.6	8.9	252	1	US-08-689-190-1
12	35.6	8.9	252	2	US-08-733-446-23
13	35.6	8.9	252	2	US-08-733-446-56
14	35.6	8.9	252	2	US-08-733-446-57
15	35.6	8.9	252	2	US-08-733-446-58
16	35.6	8.9	252	2	US-08-733-446-62
17	35.6	8.9	252	2	US-08-835-231-34
18	35.6	8.9	252	4	US-08-108-661-34
19	35.6	8.9	253	2	US-08-733-446-29
20	35.6	8.9	254	2	US-08-733-446-30
21	35.6	8.9	256	2	US-08-733-446-31
22	35.6	8.9	263	1	US-08-689-190-3
23	35.6	8.9	263	2	US-08-733-446-32
24	35.6	8.9	265	1	US-08-689-190-4
25	35.6	8.9	265	2	US-08-733-446-33
26	35.2	8.8	7218	1	US-08-232-463-14
27	33.8	8.4	183	4	US-09-134-001C-981

c	28	32.4	8.1	51259	3	US-08-781-891-209
	29	32	8.0	812	1	US-07-612-674-3
	30	32	8.0	1696	1	US-07-612-674-1
c	31	31.6	7.9	5855	1	US-08-592-214A-20
c	32	31.6	7.9	5855	3	US-09-149-976-20
	33	31	7.7	29793	4	US-09-302-812-38
	34	31	7.7	29793	4	US-09-511-477-38
	35	31	7.7	29793	4	US-09-511-507-38
c	36	30.6	7.6	432	4	US-08-905-223-134
c	37	30.4	7.6	2606	2	US-08-568-459A-7
c	38	30.4	7.6	2606	2	US-08-487-826B-7
c	39	30.4	7.6	2606	4	US-09-210-288-7
c	40	30	7.5	1669	4	US-09-026-039-1
c	41	29.6	7.4	931	4	US-09-453-702B-112
	42	29.6	7.4	1609	4	US-09-312-285-1
	43	29.6	7.4	1609	4	US-09-728-764-1
	44	29.6	7.4	14672	4	US-08-961-527-111
	45	29.4	7.3	1557	4	US-09-134-001C-1614

ALIGNMENTS

RESULT 1
US-08-733-446-19
Sequence 19, Application US/08733446
Patent No. 5856138
GENERAL INFORMATION:
APPLICANT: FUKUDA, Tsunehiko
TITLE OF INVENTION: HUMAN PARATHYROID HORMONE MUTAINS AND
TITLE OF INVENTION: PRODUCTION THEREOF
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,446
FILING DATE: 18-Oct-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/402,970
FILING DATE:
APPLICATION NUMBER: US/07/926,787
ATTORNEY/AGENT INFORMATION:
NAME: NEUMER, George W
REGISTRATION NUMBER: 26964
REFERENCE/DOCKET NUMBER: 42025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..234
IDENTIFICATION METHOD: E


```

CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,446
FILING DATE: 18-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/402,970
FILING DATE:
APPLICATION NUMBER: US/07/926,787
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NEUNER, George W
REGISTRATION NUMBER: 26964
REFERENCE/DOCKET NUMBER: 42025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 248 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
ANTI-SENSE: NO
US-08-733-446-26

Query Match
Best Local Similarity 8.9%; Score 35.6; DB 2; Length 248;
Pred. No. 0.037;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 65 GAATTGTGTCGTCGACCCCTATTATTGTCAGATGTCGACATGGAGCTTATATTA 124
DB 67 GAATTCGAGATGTGCACAAATTTGTCCTTAGTGCCCCCATGGCTCCTCGTAGGC 126
QY 125 AGCTTCGATGAGAAGAGAAGAGATGTCAGAGCTTTCACAGACCCGAGAAAACAA 184
DB 127 TGGTCCCAAGACAGACGTAAGAAAGACAAATGCTTAGTGAGAGCCATGAAAAATC 186
QY 185 CC 186
DB 187 CC 188

RESULT 9
US-08-733-446-27/c
Sequence 27, Application US/08733446
Patent No. 5856138
GENERAL INFORMATION:
APPLICANT: FUKUDA, Tsunehiko
TITLE OF INVENTION: HUMAN PARATHYROID HORMONE MUTAINS AND
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US

```

```

ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,446
FILING DATE: 18-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/402,970
FILING DATE:
APPLICATION NUMBER: US/07/926,787
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NEUNER, George W
REGISTRATION NUMBER: 26964
REFERENCE/DOCKET NUMBER: 42025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 250 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid, synthetic DNA
ANTI-SENSE: YES
US-08-733-446-27

Query Match
Best Local Similarity 8.9%; Score 35.6; DB 2; Length 250;
Pred. No. 0.037;
Matches 68; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 65 GAATTGTGTCGTCGACCCCTATTATTGTCAGATGTCGACATGGAGCTTATATTA 124
DB 186 GAATTCGAGATGTGCACAAATTTGTCCTTAGTGCCCCCATGGCTCCTCGTAGGC 127
QY 125 AGCTTCGATGAGAAGAGAAGAGATGTCAGAGCTTTCACAGACCCGAGAAAACAA 184
DB 126 TGGTCCCAAGACAGACGTAAGAAAGACAAATGCTTAGTGAGAGCCATGAAAAATC 67
QY 185 CC 186
DB 66 CC 65

RESULT 10
US-08-733-446-28
Sequence 28, Application US/08733446
Patent No. 5856138
GENERAL INFORMATION:
APPLICANT: FUKUDA, Tsunehiko
TITLE OF INVENTION: HUMAN PARATHYROID HORMONE MUTAINS AND
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN, DIKE, BRONSTEIN, ROBERTS &
ADDRESSEE: CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/733,446
FILING DATE: 18-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

RESULT 14
 US-08-733-446-57
 Sequence 57, Application US/08733446
 Patent No. 5856138
 GENERAL INFORMATION:
 APPLICANT: FUKUDA, Tsunehiko
 TITLE OF INVENTION: HUMAN PARATHYROID HORMONE MUTEINS AND
 TITLE OF INVENTION: PRODUCTION THEREOF
 NUMBER OF SEQUENCES: 62
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS &
 ADDRESSEE: CUSHMAN
 STREET: 130 Water Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: US
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/733,446
 FILING DATE: 18-Oct-1996
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/402,970
 FILING DATE:
 APPLICATION NUMBER: US/07/926,787
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: NEUNER, George W

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 13:22:11 ; Search time 84 Seconds
(Without alignments)
2069,842 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccggcgccggctccaccagct.....gycataacttcctgtgtcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 381593 seqs, 216252194 residues

Total number of hits satisfying chosen parameters: 763186

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PC7_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PC7US_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.4	9.1	2553	10 US-09-835-995-1	Sequence 1, Appl1
C 2	34.8	8.7	1984	10 US-09-854-133-390	Sequence 390, App
C 3	34.8	8.7	1984	10 US-09-738-973-390	Sequence 618, App
C 4	33.4	8.3	989	10 US-09-070-927A-618	Sequence 1449, Ap
C 5	32.2	8.0	642	10 US-09-764-847-1449	Sequence 1450, Ap
C 6	32.2	8.0	642	10 US-09-764-847-1450	Sequence 1451, Ap
C 7	32.2	8.0	642	10 US-09-764-847-1451	Sequence 1452, Ap
C 8	32.2	8.0	642	10 US-09-764-847-1452	Sequence 1450, Ap
C 9	32.2	8.0	855	10 US-09-815-242-4500	Sequence 8275, Ap
C 10	32.2	8.0	861	10 US-09-815-242-8275	Sequence 8275, Ap
C 11	32	8.0	180216	10 US-09-835-232-6	Sequence 6, Appl1
C 12	31.6	7.9	1984	10 US-09-864-761-4541	Sequence 4541, Ap
C 13	31.2	7.8	401	9 US-09-946-807-423	Sequence 423, App
C 14	31.2	7.8	401	9 US-09-946-807-424	Sequence 425, App
C 15	31.2	7.8	401	9 US-09-946-807-425	Sequence 426, App
C 16	31.2	7.8	401	9 US-09-946-807-426	Sequence 423, App
C 17	31.2	7.8	401	10 US-09-795-668-423	Sequence 424, App
C 18	31.2	7.8	401	10 US-09-795-668-424	Sequence 425, App
C 19	31.2	7.8	401	10 US-09-795-668-425	Sequence 425, App

C 20	31.2	7.8	401	10 US-09-795-668-426	Sequence 426, App
C 21	31.2	7.8	401	10 US-09-795-686-423	Sequence 423, App
C 22	31.2	7.8	401	10 US-09-795-686-424	Sequence 425, App
C 23	31.2	7.8	401	10 US-09-795-686-425	Sequence 426, App
C 24	31.2	7.8	401	10 US-09-795-686-426	Sequence 264, App
C 25	31.2	7.8	401	10 US-09-764-847-264	Sequence 4361, Ap
C 26	31	7.7	411	10 US-09-983-965-4361	Sequence 38, Appl
C 27	31	7.7	29793	10 US-09-973-451-38	Sequence 12123, A
C 28	30.8	7.7	434	10 US-09-960-355-12123	Sequence 11, Appl
C 29	30.8	7.7	2141	10 US-09-841-786-11	Sequence 8, Appl
C 30	30.8	7.7	9726	10 US-09-841-786-8	Sequence 15, Appl
C 31	30.8	7.7	11130	10 US-09-841-786-15	Sequence 8879, Ap
C 32	30.6	7.6	861	10 US-09-815-242-8879	Sequence 9010, Ap
C 33	30.6	7.6	861	10 US-09-815-242-9010	Sequence 276, App
C 34	30.6	7.6	2594	10 US-09-925-300-276	Sequence 3, Appl
C 35	30.6	7.6	3030	10 US-09-906-453-3	Sequence 1, Appl
C 36	30.6	7.6	1503841	9 US-09-946-807-1	Sequence 1, Appl
C 37	30.6	7.6	1503841	10 US-09-795-668-1	Sequence 1, Appl
C 38	30.6	7.6	1503841	10 US-09-795-668-1	Sequence 3462, Ap
C 39	30.4	7.6	388	10 US-09-983-965-3462	Sequence 7, Appl
C 40	30.4	7.6	2606	9 US-10-153-273-7	Sequence 105, Ap
C 41	30.4	7.6	2660	10 US-09-833-381-1015	Sequence 7539, Ap
C 42	30.2	7.5	355	10 US-09-960-352-7539	Sequence 3151, Ap
C 43	30.2	7.5	722	9 US-09-938-842A-3151	Sequence 286, App
C 44	30	7.5	374	10 US-09-895-828-286	Sequence 3362, Ap
C 45	29.8	7.4	2000	9 US-09-938-842A-3362	

ALIGNMENTS

RESULT 1
US-09-835-995-1/c
; Sequence 1, Application US/09835995
; Patent No. US20020052021A1
; GENERAL INFORMATION:
; APPLICANT: Soderberg, Charlotte
; TITLE OF INVENTION: No. US20020052021A1el G Protein Coupled Receptors
; FILE REFERENCE: 00146reg.US
; CURRENT APPLICATION NUMBER: US/09/835,995
; PRIOR FILING DATE: 2001-04-17
; PRIOR APPLICATION NUMBER: 60/198,090
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2553
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-835-995-1

Query Match 9.1%; Score 36.4; DB 10; Length 2553;
Best Local Similarity 54.5%; Pred. No. 0.15;
Matches 73; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 151 ATGTCAGAGTCTTACAGAGACCGAGAAACACACGATTCAGGGCCAGGCGCTTGA 210
Db ATTCACACCTCTGGAGAGATGACAGAAAGTACACAGACGACGACGACGCTGGCA 769
QY 211 ATCCATATTCCTTTCAGCTCAGTGAGCAATTTTGAAGCAATATTTAAATCTGAA 270
Db CAGAGACATTCACCTTTGATTTGAGGTAGACATTTAGCTGATGACATTTTGTAAATTTGCA 709
QY 271 GGGGTAAGAACAA 284
Db TATGTAACACACTA 695
RESULT 2
US-09-854-133-390/c
; Sequence 390, Application US/09854133
; Publication No. US20020183499A1

```
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-854-133-390
```

```
Query Match 8.7%; Score 34.8; DB 9; Length 1984;
Best Local Similarity 53.7%; Pred. No. 0.41;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
QY 151 ATGTCAAGTCTTCACAGAGACCGAGAAACACACAGATTCCAGGGCCACAGGCTTGA 210
   || |||| ||| || || || || || || || || || || || || || || || ||
DB 908 ATACACAGCTCTGAGAGACTACAGAGAAATGACTACAGAGCCACGATGACAGTCTGCA 849
QY 211 ATCCCATATTTCTTGCAGTTCAGTGACAAATTTAGAGGCAATTTAAATCTGAA 270
   || || || || || || || || || || || || || || || || || || || ||
DB 848 CAGAGACATACCTTTGATTCGAGTAGACATTAGCTGATGACATTTTGTATTTTGA 789
QY 271 GGGTTAAGACAA 284
   || || || || || || || || || || || || || || || || || || || ||
DB 788 TATGTAACACACTA 775
```

```
RESULT 3
US-09-738-973-390/c
; Sequence 390, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Elliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 390
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-738-973-390
```

```
Query Match 8.7%; Score 34.8; DB 10; Length 1984;
Best Local Similarity 53.7%; Pred. No. 0.41;
Matches 72; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
```

```
QY 151 ATGTCAAGTCTTCACAGAGACCGAGAAACACACAGATTCCAGGGCCACAGGCTTGA 210
   || |||| ||| || || || || || || || || || || || || || || || ||
DB 908 ATACACAGCTCTGAGAGACTACAGAGAAATGACTACAGAGCCACGATGACAGTCTGCA 849
```

```
QY 211 ATCCCATATTTCTTGCAGTTCAGTGACAAATTTAGAGGCAATTTAAATCTGAA 270
   || || || || || || || || || || || || || || || || || || || ||
DB 848 CAGAGACATACCTTTGATTCGAGTAGACATTAGCTGATGACATTTTGTATTTTGA 789
QY 271 GGGTTAAGACAA 284
   || || || || || || || || || || || || || || || || || || || ||
DB 788 TATGTAACACACTA 775
```

```
RESULT 4
US-09-070-927A-618/c
; Sequence 618, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillon
; Steven Barash
```

```
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
```

```
; ZIP: 20850
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <unknown>
```

```
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
```

```
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
```

```
; TELEFAX: (301) 309-8512
```

```
; INFORMATION FOR SEQ ID NO: 618:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 989 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
```

```
; SEQUENCE DESCRIPTION: SEQ ID NO: 618:
US-09-070-927A-618
```

```
Query Match 8.3%; Score 33.4; DB 10; Length 989;
Best Local Similarity 62.7%; Pred. No. 0.79;
Matches 52; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
```

```
QY 238 GACATTTTAGAGGCAATTTAAATCTGAGAGGCTAAAGAAACAATCAGATGTGTTA 297
   || || || || || || || || || || || || || || || || || || || ||
DB 655 GAAATATTAAGAGAGCAATTTTAAATTTGCCAATGATAATTAATTAAGCTTTTGTA 596
QY 298 ATTGTGAGTTCAGATCATAT 320
   || || || || || || || || || || || || || || || || || || || ||
DB 595 TTTTGTGTGTATAGAAATAAT 573
```

```
RESULT 5
```

```
US-09-764-847-1449
; Sequence 1449, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1449
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1449

Query Match
Best Local Similarity 8.0%; Score 32.2; DB 10; Length 642;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 277 AAGAACAATCAGATGCTTAATGTTGAGGTTGAGATCATTAATGCCATCTGATGCAATTC 336
Db 311 AAAAATTGTCACATGCTGTAAGTGTGTGAGAAAAATAAATACACTTAATGAATTT 370
OY 337 CATTTCTGCTTCTTT 353
Db 371 CATTTTCTTCTCTGT 387

RESULT 6
US-09-764-847-1450
; Sequence 1450, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1450
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1450

Query Match
Best Local Similarity 8.0%; Score 32.2; DB 10; Length 642;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 277 AAGAACAATCAGATGCTTAATGTTGAGGTTGAGATCATTAATGCCATCTGATGCAATTC 336
Db 311 AAAAATTGTCACATGCTGTAAGTGTGTGAGAAAAATAAATACACTTAATGAATTT 370
OY 337 CATTTCTGCTTCTTT 353
Db 371 CATTTTCTTCTCTGT 387

RESULT 7
US-09-764-847-1451
; Sequence 1451, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
```

```
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1451
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1451

Query Match
Best Local Similarity 8.0%; Score 32.2; DB 10; Length 642;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 277 AAGAACAATCAGATGCTTAATGTTGAGGTTGAGATCATTAATGCCATCTGATGCAATTC 336
Db 311 AAAAATTGTCACATGCTGTAAGTGTGTGAGAAAAATAAATACACTTAATGAATTT 370
OY 337 CATTTCTGCTTCTTT 353
Db 371 CATTTTCTTCTCTGT 387

RESULT 8
US-09-764-847-1452
; Sequence 1452, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1452
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1452

Query Match
Best Local Similarity 8.0%; Score 32.2; DB 10; Length 642;
Matches 49; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 277 AAGAACAATCAGATGCTTAATGTTGAGGTTGAGATCATTAATGCCATCTGATGCAATTC 336
Db 311 AAAAATTGTCACATGCTGTAAGTGTGTGAGAAAAATAAATACACTTAATGAATTT 370
OY 337 CATTTCTGCTTCTTT 353
Db 371 CATTTTCTTCTCTGT 387

RESULT 9
US-09-815-242-4500/c
; Sequence 4500, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
```

```
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 4500
; LENGTH: 855
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4500
```

```
Query Match      8.0%; Score 32.2; DB 10; Length 855;
Best Local Similarity 51.0%; Pred. No. 1.8;
Matches 76; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
```

```
OY 206 TTGAATCCCATATTTCTTGTACTGACGTGACAAATTTTGAAGCAATATTTAAATC 265
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 711 TTGGTTTTCAGTATTACGTTAATTTTACTGACCAATGGAATGCTTTTGGATATC 652
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 266 TCGAAGGGGTAAAGAACAAATCAGATGCTTTAATTTGTGAGGTTCCAGATCAATGCCAT 325
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 651 TTAGTCGGGATACAGTACCGCGTGAATGTAATGTAACCTGTAAACCTGTAAACCGAT 592
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 326 CTGATGCATTCATTTCTGTTCTTTG 354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 591 TTCTTCATTTCTTTAAATCTAATTTTG 563
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 10

US-09-815-242-8275/C

; Sequence 8275, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Karl L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8275

```
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(861)
US-09-815-242-8275
```

```
Query Match      8.0%; Score 32.2; DB 10; Length 861;
Best Local Similarity 51.0%; Pred. No. 1.8;
Matches 76; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
```

```
OY 206 TTGAATCCCATATTTCTTGTACTGACGTGACAAATTTTGAAGCAATATTTAAATC 265
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 711 TTGGTTTTCAGTATTACGTTAATTTTACTGACCAATGGAATGCTTTTGGATATC 652
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 266 TCGAAGGGGTAAAGAACAAATCAGATGCTTTAATTTGTGAGGTTCCAGATCAATGCCAT 325
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 651 TTAGTCGGGATACAGTACCGCGTGAATGTAATGTAACCTGTAAACCTGTAAACCGAT 592
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 326 CTGATGCATTCATTTCTGTTCTTTG 354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 591 TTCTTCATTTCTTTAAATCTAATTTTG 563
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 11

US-09-835-232-6

; Sequence 6, Application US/09835232

; Patent No. US20020098489A1

; GENERAL INFORMATION:

; APPLICANT: Leder, Philip

; APPLICANT: Leader, Benjamin

; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES

; FILE REFERENCE: 00383/052002

; CURRENT APPLICATION NUMBER: US/09/835,232

; PRIOR FILING DATE: 2001-04-12

; PRIOR APPLICATION NUMBER: US 60/196,811

; PRIOR FILING DATE: 2000-04-13

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 6

; LENGTH: 180216

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: (1)...(180216)

; OTHER INFORMATION: n = A,T,C or G

US-09-835-232-6

```
Query Match      8.0%; Score 32; DB 10; Length 180216;
Best Local Similarity 51.4%; Pred. No. 38;
Matches 74; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
```

```
OY 131 GATGAGAAGAGAAAGAGATGTCAGGCTTTCACAGACAGACAGAGAAACACACAGAT 190
    || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 22235 GAAAAAAGAAAAAAGAAAGAAAGAGAGAGAGGAGAAAAAAGACAAACACATAT 22234
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 191 TCCAGGGCCAGAGGCTTGAATCCATATTTCTTGTACTGACGTGACAAATTTAGAA 250
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 22295 TCCATTAAGCATTTGATTGTATACAAATATGTTCTCAGGTTCTGTATGCAATATTAAG 22354
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 251 GGCATATTTTAATCTGGAAGGG 274
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 22355 TGTTTTCTTAAACAGAGGCTTGG 22378
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

RESULT 12

US-09-864-761-4541

; Sequence 4541, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

```

; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4541
; LENGTH: 1984
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002094.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 31
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 31
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 36
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
; US-09-864-761-4541

Query Match      7.9%; Score 31.6; DB 10; Length 1984;
Best Local Similarity 48.4%; Pred. No. 4.4;
Matches 88; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```

```

OY 207 TGAATCCATATTTCTTGACTTCAGGTGAGCAATTTAGAGCAATTTAAATCT 266
DB 1712 TTGAATTTCTATTATAGATATACATTTCTTCTTTGAATGCTATTTCTAGT 1771

```

```

OY 267 GGAAGGGTAAGACATCATGATGTTAATTTGGAGTTCCAGATCAATTAATGCATC 326
DB 1772 TTATGATATTTGAAACATCTTTGATTTATCTTTTAATTAAGTTCACTGCTCTT 1831
OY 327 TGATGATTCATTTCTGTTCTTTTGAACACAGTACAAAAACCAAGGGAGGGCA 386
DB 1832 TCCAGCAATATCTTTCTTTCTTTGTCAGTAGTAAACAAATCAAGCAAGCA 1891
OY 387 TA 388
DB 1892 TA 1893

RESULT 13
US-09-946-807-423/c
; Sequence 423, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Gulcher, Jeffrey R.
; APPLICANT: Steinthorsdottir, Valgerdur
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 423
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-946-807-423

Query Match      7.8%; Score 31.2; DB 9; Length 401;
Best Local Similarity 47.6%; Pred. No. 2.4;
Matches 81; Conservative 3; Mismatches 86; Indels 0; Gaps 0;

```

```

OY 178 AAAACAACAGATTCAGAGGCAAGGGCTTGAATCCATATTTCTTGACTTCAGTG 237
DB 301 ARAAATTTGAGGACACTTCTAGTTCTTTGGCCCTTATTTCCATTTCCAAAGRA 242
OY 238 GACATTTTGAAGGCAATTTTAATCTGGAAGGGTAAAGAAACAATTCAGATGCTT 297
DB 241 AGCCAAAGTTAAGTCTACTCAAAATGTGCGGCAATTTMACACAAATGATTTGATA 182
OY 298 ATTGTGAGGTTTCAGATCATTAATCCATCTGATGCATTCATTTCTGCTT 347
DB 181 ACAAGTGAATTTGACAAAGCATGTCTTTTAATCATGATGTTTCTT 132

RESULT 14
US-09-946-807-424/c
; Sequence 424, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinthorsdottir, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; CURRENT FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 424

```

```

; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-946-807-424

```

Query Match	7.88;	Score 31.2;	DB 9;	Length 401;
Best Local Similarity	47.68;	Pred. No. 2.4;		
Matches 81; Conservative	3;	Mismatches 86;	Indels 0;	Gaps 0;

[illegible]

RESULT 15

```

US-09-946-807-425/c
: Sequence 425, Application US/0946807
: Patent No. US0020165144A1
:
: GENERAL INFORMATION:
: APPLICANT: Stefansson, Hreinn
: APPLICANT: Steinhorsdottir, Valgerdur
: APPLICANT: Gulicher, Jeffrey R.
: TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
: FILE REFERENCE: 2345, 2004-001
: CURRENT APPLICATION NUMBER: US/09/946,807
: CURRENT FILING DATE: 2001-09-05
: PRIOR APPLICATION NUMBER: US/09/795,668
: PRIOR FILING DATE: 2001-02-28
: PRIOR APPLICATION NUMBER: US 09/515,716
: PRIOR FILING DATE: 2000-02-28
: NUMBER OF SEQ ID NOS: 1531
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 425
: LENGTH: 401
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-946-807-425

```

	Query Match	7.8%	Score 31.2;	DB 9;	Length 401;
	Best Local Similarity	47.6%;	Pred. No. 2.4;		
	Matches	81;	Conservative	3;	Mismatches 86; Indels 0; Gaps 0.
OY	178	AAACCAACCAGATTCCAGGGCCGGAAGCTTGGAAATCCCATATTTTCCTTGACTCAGGTG	237		
	: :				
Db	259	ARAAAATTGAGAGCGAGCTCTAGTTCTTTGGGCCCTTCATTTCTTATTTCCAGGRA	200		
OY	238	GACATTTTAGAAGGCATATTTAAATCTGGAAGGGGTAAGACAACAATCAGATGGTGTTA	297		
	:	:	:		
Db	199	AGCCAAGTTAAAGTCTACTCARAATGTGGCGGCATTTCACAAATAGATGTGATA	140		
OY	298	ATTGTGAGTTTCAGATCATATGCGCATCTGATGATTCATTCATTCMGGT	347		
Db	139	ACAAGTGAATTTGACNAAGCATGTGCTTTTATATCATGTAATTTTTCCCT	90		

Search completed: January 2, 2003, 15:06:42
Job time : 129 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:27:57 : Search time 3761 Seconds

(without alignments)
2687.392 Million cell updates/sec

Title: US-09-850-147-1

Perfect score: 402

Sequence: 1 ccggcgccggcgtccaccagct.....ggcataactctctgtgtcc 402

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main:*

- 1: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
- 2: /cgn2_6/ptodata/1/pna/US086_COMB.seq.*
- 3: /cgn2_6/ptodata/1/pna/US087_COMB.seq.*
- 4: /cgn2_6/ptodata/1/pna/US088_COMB.seq.*
- 5: /cgn2_6/ptodata/1/pna/US089_COMB.seq.*
- 6: /cgn2_6/ptodata/1/pna/US090_COMB.seq.*
- 7: /cgn2_6/ptodata/1/pna/US091_COMB.seq.*
- 8: /cgn2_6/ptodata/1/pna/US092_COMB.seq.*
- 9: /cgn2_6/ptodata/1/pna/US093_COMB.seq.*
- 10: /cgn2_6/ptodata/1/pna/US094_COMB.seq.*
- 11: /cgn2_6/ptodata/1/pna/US095_COMB.seq.*
- 12: /cgn2_6/ptodata/1/pna/US096_COMB.seq.*
- 13: /cgn2_6/ptodata/1/pna/US097_COMB.seq.*
- 14: /cgn2_6/ptodata/1/pna/US098_COMB.seq.*
- 15: /cgn2_6/ptodata/1/pna/US099_COMB.seq.*
- 16: /cgn2_6/ptodata/1/pna/US100_COMB.seq.*
- 17: /cgn2_6/ptodata/1/pna/US101_COMB.seq.*
- 18: /cgn2_6/ptodata/1/pna/US102_COMB.seq.*
- 19: /cgn2_6/ptodata/1/pna/US103_COMB.seq.*
- 20: /cgn2_6/ptodata/1/pna/US104_COMB.seq.*
- 21: /cgn2_6/ptodata/1/pna/US105_COMB.seq.*
- 22: /cgn2_6/ptodata/1/pna/US106_COMB.seq.*
- 23: /cgn2_6/ptodata/1/pna/US107_COMB.seq.*
- 24: /cgn2_6/ptodata/1/pna/US108_COMB.seq.*
- 25: /cgn2_6/ptodata/1/pna/US109_COMB.seq.*
- 26: /cgn2_6/ptodata/1/pna/US110_COMB.seq.*
- 27: /cgn2_6/ptodata/1/pna/US111_COMB.seq.*
- 28: /cgn2_6/ptodata/1/pna/US112_COMB.seq.*
- 29: /cgn2_6/ptodata/1/pna/US113_COMB.seq.*
- 30: /cgn2_6/ptodata/1/pna/US114_COMB.seq.*
- 31: /cgn2_6/ptodata/1/pna/US115_COMB.seq.*
- 32: /cgn2_6/ptodata/1/pna/US116_COMB.seq.*
- 33: /cgn2_6/ptodata/1/pna/US117_COMB.seq.*
- 34: /cgn2_6/ptodata/1/pna/US118_COMB.seq.*
- 35: /cgn2_6/ptodata/1/pna/US119_COMB.seq.*
- 36: /cgn2_6/ptodata/1/pna/US120_COMB.seq.*
- 37: /cgn2_6/ptodata/1/pna/US121_COMB.seq.*
- 38: /cgn2_6/ptodata/1/pna/US122_COMB.seq.*
- 39: /cgn2_6/ptodata/1/pna/US123_COMB.seq.*
- 40: /cgn2_6/ptodata/1/pna/US124_COMB.seq.*
- 41: /cgn2_6/ptodata/1/pna/US125_COMB.seq.*
- 42: /cgn2_6/ptodata/1/pna/US126_COMB.seq.*
- 43: /cgn2_6/ptodata/1/pna/US127_COMB.seq.*

Result No.	Score	Query Match	Length	DB ID	Description
1	402	100.0	402	25	US-09-654-617-457554
2	402	100.0	402	27	US-09-684-016-457554
3	402	100.0	402	32	US-09-850-147-1
4	402	100.0	402	64	US-60-202-213-1
5	266.8	66.4	426	32	US-09-850-147-40
6	266.8	66.4	426	64	US-60-202-213-40
7	266.8	66.4	426	25	US-09-654-617-451847
8	266.8	66.4	426	32	US-09-684-016-451847
9	266.8	66.4	426	27	US-09-850-147-1092
10	266.8	66.4	426	32	US-60-202-213-1089
11	255	63.4	385	64	US-09-850-147-10
12	255	63.4	385	64	US-60-202-213-10
13	223.6	55.6	1446	28	US-09-708-427-52180
14	223.6	55.6	1446	19	US-09-513-966A-50583
15	221.4	55.1	1694	25	US-09-654-617-262628
16	221.4	55.1	1694	27	US-09-684-016-262628
17	220.6	54.9	570	33	US-09-865-439A-116358
18	220	54.7	484	33	US-09-865-419A-15731
19	220	54.7	484	64	US-60-208-063-2535
20	220	54.7	545	33	US-09-873-402A-41873
21	220	54.7	550	33	US-09-865-419A-51912

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```

c 22 220 54.7 559 33 US-09-873-402A-54657 Sequence 54657, A
c 23 220 54.7 559 64 US-60-209-830-46016 Sequence 46016, A
c 24 220 54.7 995 42 US-10-219-999-1500 Sequence 1500, Ap
c 25 220 54.7 1325 42 US-10-219-999-18314 Sequence 18314, A
c 26 217 54.0 472 33 US-09-873-402A-11200 Sequence 11200, A
c 27 217 54.0 472 64 US-60-209-830-11200 Sequence 11200, A
c 28 213.6 53.1 450 64 US-60-207-458-54563 Sequence 54563, A
c 29 213 53.0 535 33 US-09-865-439A-115458 Sequence 115458, A
c 30 213 53.0 1480 42 US-10-219-999-22880 Sequence 22880, A
c 31 208.2 51.8 1231 42 US-10-219-999-20024 Sequence 20024, A
c 32 205.6 51.1 533 33 US-09-865-439A-80117 Sequence 80117, A
c 33 205.6 51.1 553 33 US-60-207-458-124259 Sequence 124259, A
c 34 202.4 50.3 479 33 US-09-865-419A-42700 Sequence 42700, A
c 35 201.6 50.1 393 22 US-09-565-306-82924 Sequence 82924, A
c 36 196.4 48.9 397 64 US-60-207-458-43952 Sequence 43952, A
c 37 195 48.5 388 33 US-09-865-439A-18142 Sequence 18142, A
c 38 193 48.0 501 64 US-60-207-458-25263 Sequence 25263, A
c 39 192.6 47.9 411 33 US-09-865-419A-9944 Sequence 9944, A
c 40 184.2 45.8 351 33 US-09-865-439A-4361 Sequence 4361, Ap
c 41 181.8 45.2 500 33 US-09-865-439A-93679 Sequence 93679, A
c 42 181.8 44.5 500 64 US-60-207-458-137694 Sequence 137694, A
c 43 179 44.5 396 33 US-09-865-439A-89549 Sequence 89549, A
c 44 179 44.5 396 64 US-60-207-458-133564 Sequence 133564, A
c 45 173.4 43.1 520 30 US-09-795-301-466 Sequence 466, App

```

ALIGNMENTS

```

RESULT 1
US-09-654-617-457554
; Sequence 457554, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654, 617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 457554
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
US-09-654-617-457554

Query Match 100.0%; Score 402; DB 25; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.3e-109;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CCGGCCCGGCTCACACCTGGCCAGTGAAGGCTTAAGACATGTGCCAGAGAAGGG 60
DB 1 CCGGCCCGGCTCACACCTGGCCAGTGAAGGCTTAAGACATGTGCCAGAGAAGGG 60
QY 61 CTTGAACCTTGTTGTTGCTGCACCTTAATTTGCTCAGATGTGCAATGAGCTTAT 120
DB 61 CTTGAACCTTGTTGTTGCTGCACCTTAATTTGCTCAGATGTGCAATGAGCTTAT 120
QY 121 ATTAAGCTTCAGATGAGAAGAAAGAGATGTCAGTCTTCACAGACCGAGAAA 180
DB 121 ATTAAGCTTCAGATGAGAAGAAAGAGATGTCAGTCTTCACAGACCGAGAAA 180
QY 181 ACAACACATTCACAGGCGCAAGGCTTGAATCCATATTTCTTGAATCAGGTGAC 240
DB 181 ACAACACATTCACAGGCGCAAGGCTTGAATCCATATTTCTTGAATCAGGTGAC 240
QY 241 AATTTAGAGCAATATTTAAATCTGGAAGGGTAAAGAACATCAATGTTAAT 300
DB 241 AATTTAGAGCAATATTTAAATCTGGAAGGGTAAAGAACATCAATGTTAAT 300
QY 301 GTGAGGTTGATGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 GTGAGGTTGATGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 360

```

```

QY 361 AGTACAAAAAACCAACGAGGAGGATTAATCTCTGTGTC 402
DB 361 AGTACAAAAAACCAACGAGGAGGATTAATCTCTGTGTC 402

```

```

RESULT 2
US-09-684-016-457554
; Sequence 457554, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jindong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684, 016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654, 617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 457554
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
US-09-684-016-457554

```

```

Query Match 100.0%; Score 402; DB 27; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.3e-109;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 CCGGCCCGGCTCACACCTGGCCAGTGAAGGCTTAAGACATGTGCCAGAGAAGGG 60
DB 1 CCGGCCCGGCTCACACCTGGCCAGTGAAGGCTTAAGACATGTGCCAGAGAAGGG 60
QY 61 CTTGAACCTTGTTGTTGCTGCACCTTAATTTGCTCAGATGTGCAATGAGCTTAT 120
DB 61 CTTGAACCTTGTTGTTGCTGCACCTTAATTTGCTCAGATGTGCAATGAGCTTAT 120
QY 121 ATTAAGCTTCAGATGAGAAGAAAGAGATGTCAGTCTTCACAGACCGAGAAA 180
DB 121 ATTAAGCTTCAGATGAGAAGAAAGAGATGTCAGTCTTCACAGACCGAGAAA 180
QY 181 ACAACACATTCACAGGCGCAAGGCTTGAATCCATATTTCTTGAATCAGGTGAC 240
DB 181 ACAACACATTCACAGGCGCAAGGCTTGAATCCATATTTCTTGAATCAGGTGAC 240
QY 241 AATTTAGAGCAATATTTAAATCTGGAAGGGTAAAGAACATCAATGTTAAT 300
DB 241 AATTTAGAGCAATATTTAAATCTGGAAGGGTAAAGAACATCAATGTTAAT 300
QY 301 GTGAGGTTGATGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 GTGAGGTTGATGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AGTACAAAAAACCAACGAGGAGGATTAATCTCTGTGTC 402
DB 361 AGTACAAAAAACCAACGAGGAGGATTAATCTCTGTGTC 402

```

```

RESULT 3
US-09-850-147-1
; Sequence 1, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerston, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850, 147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202, 213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654, 617
; PRIOR FILING DATE: 2000-09-05

```



```
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-09-850-147-1

Query Match      100.0%; Score 402; DB 32; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.3e-109;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 60
D 1 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 60
QY 61 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 120
D 61 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 120
QY 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
D 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
QY 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
D 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
QY 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
D 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
QY 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
D 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
QY 361 AGTACAAAACCAACGAGGGGCAATACCTCTGTTCC 402
D 361 AGTACAAAACCAACGAGGGGCAATACCTCTGTTCC 402

RESULT 4
US-60-202-213-1
; Sequence 1, Application US/60202213
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; TITLE OF INVENTION: Edgerton, Michael D.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; FILE REFERENCE: 38-21(51914)A
; CURRENT APPLICATION NUMBER: US/60/202,213
; CURRENT FILING DATE: 2000-05-08
; NUMBER OF SEQ ID NOS: 17986
; SEQ ID NO 1
; LENGTH: 402
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K1-A3
US-60-202-213-1

Query Match      100.0%; Score 402; DB 64; Length 402;
Best Local Similarity 100.0%; Pred. No. 1.3e-109;
Matches 402; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 60
D 1 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 60
QY 61 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 120
D 61 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 120
QY 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
D 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
QY 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
D 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
QY 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
D 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
QY 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
D 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
```

```
QY 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
D 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
QY 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
D 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
QY 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
D 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
QY 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
D 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360

RESULT 5
US-09-850-147-40
; Sequence 40, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 40
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-09-850-147-40

Query Match      66.4%; Score 266.8; DB 32; Length 426;
Best Local Similarity 83.8%; Pred. No 5.3e-69;
Matches 337; Conservative 0; Mismatches 62; Indels 3; Gaps 3;

QY 1 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 60
D 1 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 60
QY 25 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 83
D 25 CCGGGCCGGGCTCACCGAGTGGGCCAGTGAAGGCTTAAGACATGTGCCGAGGAAGGG 83
QY 61 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 120
D 61 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 120
QY 84 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 143
D 84 CCTTGAACCTGTGTGTCTGCACCCCTAATTATGCTCAGATGTGGCAATGGAGCTTAT 143
QY 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
D 121 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 180
QY 144 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 202
D 144 ATTAAGCTTCGATGAGAAGAAAGAGAGATGTCAGAGTCTTCACAGAGACGAGGAAA 202
QY 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
D 181 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 240
QY 203 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 261
D 203 ACAACCGATTCACAGGGCCCAAGGCTTGGAATCCCATATTTCTTTCAGTGCAGTGAC 261
QY 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
D 241 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 300
QY 262 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 321
D 262 AATTTTGAAGGCAATTTTAAATCTGGAAGGGGTAAAGAACATCGATGGTTTAAAT 321
QY 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
D 301 GTGAGGTTGAGATCATTAATGCAATCTGATGCATTCATTTCTGTTTGAACACA 360
```

Db 322 TTGTAGCTTCAGATCATCTGCAATCTGATGCAATTCGTTCTGTTCTTTTGAACACCA 381
Qy 361 ACTACAAAAAACCAACGGAGGGGATTAATCTCTGTGTTC 402
Db 382 TGTACAAATAGACCATCGGAGGTGCATATCTCTGTATTC 423

RESULT 6

US-60-202-213-40
; Sequence 40, Application US/60202213
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 38-21(51914)A
; CURRENT APPLICATION NUMBER: US/60/202,213
; CURRENT FILING DATE: 2000-05-08
; NUMBER OF SEQ ID NOS: 17986
; SEQ ID NO 40
; LENGTH: 426
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K1-A3
US-60-202-213-40

Query Match 66.4%; Score 266.8; DB 64; Length 426;
Best Local Similarity 83.8%; Pred. No. 5.3e-69;
Matches 337; Conservative 0; Mismatches 62; Indels 3; Gaps 3;

Qy 1 CCGGGCCGGGCTCACCAGCTGCGCCCAAGGCTTAAGACATGCGCAGAGAGGG 60
Db 25 CCGGGCCGGGCTCACCAGCTGCGCCCAAGGCTTAAGACATGCGCAGAGAGAA-GGG 83
Qy 61 CCTGAACCTGTGTGTGTGACCCCTAATTTATTTGCTGAGATGGCAATGGAGCTTTAT 120
Db 84 CCTGTACTGTGTGTGTGTGACCCCTAATTTATTTGCTGAGATGGCAATGGAGCTTTAT 143
Qy 121 ATTAAGCTTCGATGAGAAAGAGAGAGATGTCAGAGCTTTCACAGAGAGAGAA 180
Db 144 ATTAAGCTTCGATGAGAAAGAGAGAGATGTCAGAGCTTTCACAGAGAGAGAA 202
Qy 181 ACAACAGATTCAGAGGCGCAAGGCTTGAATCCCATATTTCTTTGACTTACGTGAC 240
Db 203 ACAACAGATTCAGAGGCGCAAGGCTTGAATCCCATATTTCTTTGACTTACGTGAC 261
Qy 241 AATTTAGAGCAATATTTAATCTGGAAGGGTAAAGACAATCAGATGTTAT 300
Db 262 AATTTAGAGCAATATTTAATCTGGAAGGGTAAAGACAATCAGATGTTAT 321
Qy 301 GTGAGGTTGAGATCATATATGCAATCTGATTCATTTCTGTTCTTTTGAACACA 360
Db 322 TTGTAGTTGAGATCATATATGCAATCTGATTCATTTCTGTTCTTTTGAACACA 381
Qy 361 AGTACAAAAACCAACGGAGGGGATTAATCTCTGTGTTC 402
Db 382 TGTACAAATAGACCATCGGAGGTGCATATCTCTGTATTC 423

RESULT 7

US-09-654-617-451847
; Sequence 451847, Application US/09654617
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/654,617
; CURRENT FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 451847
; LENGTH: 456
; TYPE: DNA

; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: unsure at all n locations
US-09-654-617-451847

Query Match 66.4%; Score 266.8; DB 25; Length 456;
Best Local Similarity 83.8%; Pred. No. 5.4e-69;
Matches 337; Conservative 0; Mismatches 62; Indels 3; Gaps 3;

Qy 1 CCGGGCCGGGCTCACCAGCTGCGCCCAAGGCTTAAGACATGTCGAGAGAGGG 60
Db 25 CCGGGCCGGGCTCACCAGCTGCGCCCAAGGCTTAAGACATGTCGAGAGAGAA-GGG 83
Qy 61 CCTGAACCTGTGTGTGTGACCCCTAATTTATTTGCTGAGATGGCAATGGAGCTTTAT 120
Db 84 CCTGTACTGTGTGTGTGTGACCCCTAATTTATTTGCTGAGATGGCAATGGAGCTTTAT 143
Qy 121 ATTAAGCTTCGATGAGAAAGAGAGAGATGTCAGAGCTTTCACAGAGAGAGAA 180
Db 144 ATTAAGCTTCGATGAGAAAGAGAGAGATGTCAGAGCTTTCACAGAGAGAGAA 202
Qy 181 ACAACAGATTCAGAGGCGCAAGGCTTGAATCCCATATTTCTTTGACTTACGTGAC 240
Db 203 ACAACAGATTCAGAGGCGCAAGGCTTGAATCCCATATTTCTTTGACTTACGTGAC 261
Qy 241 AATTTAGAGCAATATTTAATCTGGAAGGGTAAAGACAATCAGATGTTAT 300
Db 262 AATTTAGAGCAATATTTAATCTGGAAGGGTAAAGACAATCAGATGTTAT 321
Qy 301 GTGAGGTTGAGATCATATATGCAATCTGATTCATTTCTGTTCTTTTGAACACA 360
Db 322 TTGTAGTTGAGATCATATATGCAATCTGATTCATTTCTGTTCTTTTGAACACA 381
Qy 361 ACTACAAAAAACCAACGGAGGGGATTAATCTCTGTGTTC 402
Db 382 TGTACAAATAGACCATCGGAGGTGCATATCTCTGTATTC 423

RESULT 8

US-09-684-016-451847
; Sequence 451847, Application US/09684016
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Liu, Jingdong
; TITLE OF INVENTION: Annotated Plant Genes
; FILE REFERENCE: 38-21(15097)D
; CURRENT APPLICATION NUMBER: US/09/684,016
; CURRENT FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; NUMBER OF SEQ ID NOS: 463173
; SEQ ID NO 451847
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(456)
; OTHER INFORMATION: unsure at all n locations
US-09-684-016-451847

Query Match 66.4%; Score 266.8; DB 27; Length 456;
Best Local Similarity 83.8%; Pred. No. 5.4e-69;
Matches 337; Conservative 0; Mismatches 62; Indels 3; Gaps 3;

Qy 1 CCGGGCCGGGCTCACCAGCTGCGCCCAAGGCTTAAGACATGTCGAGAGAGGG 60
Db 25 CCGGGCCGGGCTCACCAGCTGCGCCCAAGGCTTAAGACATGTCGAGAGAGAA-GGG 83
Qy 61 CCTGAACCTGTGTGTGTGACCCCTAATTTATTTGCTGAGATGGCAATGGAGCTTTAT 120
Db 84 CCTGTACTGTGTGTGTGTGACCCCTAATTTATTTGCTGAGATGGCAATGGAGCTTTAT 143
Qy 121 ATTAAGCTTCGATGAGAAAGAGAGAGATGTCAGAGCTTTCACAGAGAGAGAA 180

```

Db 144 ATTAAGCTGTGATGAGAGATGAAA-GAGATGTCAAGTCTTCACAGAGACAGAGAGA 202
OY 181 ACAACCGATTTCCAGAGCGCTTGGAATCCCATTTTCCCTTACCTTTAGGTGAC 240
Db 203 ACAACCGATTTCCAGAGCGCTTGGAATCCCATTTTCCCTTACCTTTAGGTGAC 261
OY 241 AATTTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGCTTAAT 300
Db 262 AATTTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGCTTAAT 321
OY 301 GTGAGGCTTCAATATTCATTCATCTGATGCAATTCATTTCTGTTTCTTTGACACA 360
Db 322 TTGTAGGCTTCAATATTCATTCATCTGATGCAATTCATTTCTGTTTCTTTGACACA 381
OY 361 AGAAGCAAAAACCAAGGAGGGGCAATCTCTGTTGCC 402
Db 382 TGTAACAATAGACCATCGGAGGTGTCAATATCTCTGTTATTC 423
```

RESULT 9

US-09-850-147-1092

```

; Sequence 1092, Application US/09850147
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(51914)B
; CURRENT APPLICATION NUMBER: US/09/850,147
; CURRENT FILING DATE: 2001-05-08
; PRIOR APPLICATION NUMBER: US 60/202,213
; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654,617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684,016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 1092
; LENGTH: 433
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)-(433)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3
US-09-850-147-1092
```

Query Match 65.7%; Score 264; DB 32; Length 433;

Best Local Similarity 84.1%; Pred. No. 3,7e-68;

Matches 333; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```

OY 7 CGGGCTCACCAAGCTGCGCCAGTGAAGCCTTAAGACATGTGCCAGAGAAAGGCGCTTGA 66
Db 1 CGGGCTCACCAAGCTGCGCCAGTGAAGCCTTAAGACATGTGCCAGAGAA-GGGCCTTGT 59
OY 67 ACTGTGTGTGTGTCACCTTAATTTATTTGTCAGATGTGGCAATGACCTTCTAATTAAG 126
Db 60 AACTGTGTGTGTCACCTTAATTTATTTGTCAGATGTGGCAATGACCTTCTAATTAAG 119
OY 127 CTTGATGAGAAGAGAGAGATGTCAAGTCTTCAAGAGACCGAGAAACCAAC 186
Db 120 CTTGATGAGAAGAGAGATGTCAAGTCTTCAAGAGACCGAGAGAAACCAAC 178
OY 187 AGATTCAGAGCGCAAGGCTTGAATCCCATTTCTTTACTTCAAGTGGACAAATTT 246
Db 179 AGATTCAGAGCGCAAGGCTTGAATCCCATTTCTTTACTTCAAGTGGACAAATTT 237
OY 247 AGAAGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGCTTAATTTGAG 306
Db 238 AGAAGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGCTTAATTTGAG 297
```

```

OY 307 GTTCAGATCATATTCGACATCTGATGCAATTCATTTCTGTTTCTTTGAACACAGTACA 366
Db 298 GTTCAGATCATATTCGACATCTGATGCAATTCATTTCTGTTTCTTTGAACACAGTACA 357
OY 367 AAAAACCAGGAGGGGCAATTAATCTCTGTTGCC 402
Db 358 ATAGACCATCGGAGGTGTCAATATCTCTGTTATTC 393
```

RESULT 10

US-60-202-213-1089

; Sequence 1089, Application US/60202213

; GENERAL INFORMATION:

; APPLICANT: Andersen, Scott E.

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WI

; TITLE OF INVENTION: PLANTS

; FILE REFERENCE: 38-21(51914)A

; CURRENT APPLICATION NUMBER: US/60/202,213

; CURRENT FILING DATE: 2000-05-08

; NUMBER OF SEQ ID NOS: 17986

; SEQ ID NO 1089

; LENGTH: 433

; TYPE: DNA

; ORGANISM: Sorghum bicolor

; OTHER INFORMATION: Clone ID: LIB3476-004-Q6-K2-A3

US-60-202-213-1089

Query Match 65.7%; Score 264; DB 64; Length 433;

Best Local Similarity 84.1%; Pred. No. 3,7e-68;

Matches 333; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```

OY 7 CGGGCTCACCAAGCTGCGCCAGTGAAGCCTTAAGACATGTGCCAGAGAAAGGCGCTTGA 66
Db 1 CGGGCTCACCAAGCTGCGCCAGTGAAGCCTTAAGACATGTGCCAGAGAA-GGGCCTTGT 59
OY 67 ACTGTGTGTGTGTCACCTTAATTTATTTGTCAGATGTGGCAATGACCTTCTAATTAAG 126
Db 60 AACTGTGTGTGTCACCTTAATTTATTTGTCAGATGTGGCAATGACCTTCTAATTAAG 119
OY 127 CTTGATGAGAAGAGAGAGATGTCAAGTCTTCAAGAGACCGAGAAACCAAC 186
Db 120 CTTGATGAGAAGAGAGATGTCAAGTCTTCAAGAGACCGAGAGAAACCAAC 178
OY 187 AGATTCAGAGCGCAAGGCTTGAATCCCATTTCTTTACTTCAAGTGGACAAATTT 246
Db 179 AGATTCAGAGCGCAAGGCTTGAATCCCATTTCTTTACTTCAAGTGGACAAATTT 237
OY 247 AGAAGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGCTTAATTTGAG 306
Db 238 AGAAGCAATATTTAAATCTGGAAGGGGTAAAGAACATCAGATGCTTAATTTGAG 297
OY 307 GTTCAGATCATATTCGACATCTGATGCAATTCATTTCTGTTTCTTTGAACACAGTACA 366
Db 298 GTTCAGATCATATTCGACATCTGATGCAATTCATTTCTGTTTCTTTGAACACAGTACA 357
OY 367 AAAAACCAGGAGGGGCAATTAATCTCTGTTGCC 402
Db 358 ATAGACCATCGGAGGTGTCAATATCTCTGTTATTC 393
```

RESULT 11

US-09-850-147-10

; Sequence 10, Application US/09850147

; GENERAL INFORMATION:

; APPLICANT: Andersen, Scott E.

; APPLICANT: Edgerton, Michael D

; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With

; TITLE OF INVENTION: Plants

; FILE REFERENCE: 38-21(51914)B

; CURRENT APPLICATION NUMBER: US/09/850,147

; CURRENT FILING DATE: 2001-05-08

; PRIOR APPLICATION NUMBER: US 60/202,213

```

; PRIOR FILING DATE: 2000-05-08
; PRIOR APPLICATION NUMBER: US 09/654, 617
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 09/684, 016
; PRIOR FILING DATE: 2000-10-10
; NUMBER OF SEQ ID NOS: 18014
; SEQ ID NO 10
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-09-850-147-10

```

```

Query Match      63.4%; Score 255; DB 32; Length 385;
Best Local Similarity 83.7%; Pred. No. 1.8e-65;
Matches 324; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```

```

QY 16 CAGCTGGCCGAGTGAAGGCTTAAGACATGTGCCAGAGAAAGGGCCTTGAACCTGTGTG 75
DB 1 CAGCTGGCCGAGGAGGCTTAAAGTACATGTTCAGAGAAAGGGCCTTGAACCTGG-GTG 59
QY 76 GTCTGCACCTTAATTTATTTGTCAGATGTGCCATGTGACCTTCTATATTAAAGCTGCATGA 135
DB 60 GTCTGCACCTTAATTTATTTGTCAGATGTGCCATGTGACCTTCTATATTAAAGCTGCATGA 119
QY 136 GAAGAGAGAAAGAGATGTCAAGGTCTTTCACAGAGACCGAGAAACACACAGATTTCAG 195
DB 120 GAAGATGGAGAA-GAGATGTCAAGTCTTTCACAGAGACCGAGAGAAACACACAGATTGCAG 178
QY 196 GCGCAGGGCTTGAATGCCATATTTCTCTTCACTCAGTGCAGCAATTTTGAAGGCA 255
DB 179 GCGCAGGAGAC-TGCAGTCCCATATTTCTCTGACTTCAGTTCAGTTCAGTTCAGTTCAGTTC 237
QY 256 TATTTAAATCTGAAGAGGGGTAAAGAACATGATGTGTTAATTTGAGAGTTTCAGATC 315
DB 238 TATTTAAATCTGAAGAGGGGTAAAGTCAATGATGTGTTAATTTGAGAGTTTCAGATC 297
QY 316 AATAATCCATGTGATGATTCATTCATTTCTGTTCTTTTGAACACAAAGTACAAAAACCAA 375
DB 298 ATAGTGCAGAGCTGATGATTCATTCATTTCTGTTCTTTTGAACACATGTACAAATAGACCAT 357
QY 376 CCGGAGGGGCTAACTTCCTGTGTCC 402
DB 358 CCGGAGGTCTCATATCTCTCTATTTCC 384

```

```

RESULT 12
US-60-202-213-10
; Sequence 10, Application US/60202213
; GENERAL INFORMATION:
; APPLICANT: Andersen, Scott E.
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS
; FILE REFERENCE: 38-21(51914)A
; CURRENT APPLICATION NUMBER: US/60/202,213
; CURRENT FILING DATE: 2000-05-08
; NUMBER OF SEQ ID NOS: 17986
; SEQ ID NO 10
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; OTHER INFORMATION: Clone ID: LIB3476-001-Q6-K2-A3
US-60-202-213-10

```

```

Query Match      63.4%; Score 255; DB 64; Length 385;
Best Local Similarity 83.7%; Pred. No. 1.8e-65;
Matches 324; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

```

```

QY 16 CAGCTGGCCGAGTGAAGGCTTAAGACATGTGCCAGAGAAAGGGCCTTGAACCTGTGTG 75
DB 1 CAGCTGGCCGAGGAGGCTTAAAGTACATGTTCAGAGAGAAAGGGCCTTGAACCTGG-GTG 59

```

```

QY 76 GTCTGCACCTTAATTTATTTGTCAGATGTGCCATGTGACCTTCTATATTAAAGCTTCATGA 135
DB 60 GTCTGCACCTTAATTTATTTGTCAGATGTGCCATGTGACCTTCTATATTAAAGCTTCATGA 119
QY 136 GAAGAGAGAAAGAGATGTCAAGGCTTTCACAGAGACCGAGAAACACACAGATTTCAG 195
DB 120 GAAGATGGAGAA-GAGATGTCAAGTCTTTCACAGAGACCGAGAGAAACACACAGATTGCAG 178
QY 196 GCGCAGGGCTTGAATGCCATATTTCTCTTCACTCAGTGCAGCAATTTTGAAGGCA 255
DB 179 GCGCAGGAGAC-TGCAGTCCCATATTTCTCTGACTTCAGTTCAGTTCAGTTCAGTTCAG 237
QY 256 TATTTAAATCTGAAGAGGGGTAAAGAACATGATGTGTTAATTTGAGAGTTTCAGATC 315
DB 238 TATTTAAATCTGAAGAGGGGTAAAGTCAATGATGTGTTAATTTGAGAGTTTCAGATC 297
QY 316 AATAATCCATGTGATGATTCATTCATTTCTGTTCTTTTGAACACAAAGTACAAAAACCAA 375
DB 298 ATAGTGCAGAGCTGATGATTCATTCATTTCTGTTCTTTTGAACACATGTACAAATAGACCAT 357
QY 376 CCGGAGGGGCTAACTTCCTGTGTCC 402
DB 358 CCGGAGGTCTCATATCTCTCTATTTCC 384

```

```

RESULT 13
US-09-708-427-52180
; Sequence 52180, Application US/09708427
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTID
; TITLE OF INVENTION: THEREBY
; FILE REFERENCE: 2750-1243P
; CURRENT APPLICATION NUMBER: US/09/708,427
; CURRENT FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 85364
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52180
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1446
; OTHER INFORMATION: any n = a, g, c, t, unknown, or other
; NAME/KEY: misc_feature
; LOCATION: 1..1446
; OTHER INFORMATION: Ceres Seq. ID 1928446
US-09-708-427-52180

```

```

Query Match      55.6%; Score 223.6; DB 28; Length 1446;
Best Local Similarity 77.0%; Pred. No. 8.2e-56;
Matches 305; Conservative 4; Mismatches 83; Indels 4; Gaps 3;

```

```

QY 4 GCGCGGGCTCACAGCTGCCAGCTGAAGGCTTAAAGACATGTCCAGAGAGAGGGCCT 63
DB 1005 GCGGGGCTCACAGCTGCCAGCTGAAGGCTTAAAGACATGTGTCCAGAGAGAGGGCCT 1064
QY 64 TGAATCTGTGTGTCTGCACCTTAATTTCTAGATGTGCAATGAGCTTCTATATT 123
DB 1065 TGTACATGTGTGTCTGCACCTTAATTTCTAGATGTGCAATGAGCTTCTATATT 1124
QY 124 AAGCTTGATGAGAAAGAGAAAGAGATGTCACAGGCTTCACAGAGACCGAGAGAAACA 183
DB 1125 AAGCTTGATGAGAAAGATGAGAAAGG-GATGTCAAGTTCTTCACAGAGACCGAGAGAAACA 1183
QY 184 ACCAGATTCAGGGCCAGAGGCTTGAATCCCATATTTCTTTGACTTCAGGTGAGCAAT 243
DB 1184 ACCAGATTCAGGGCCAGAGGAC-CGCAGTCCCATATTTCTATACATTCATTCATTAAT 1242
QY 244 TTTGAAGGCAATATTTAATCTGGAAGGGGTAAAGAACATGATGATGTTAATTTGTG 303
DB 1243 TTTGAAGGTC--TATTCGAATCTGTAAAGGCTTAAAGATCAATGATGATGTTAATTTGTT 1300

```

```

RESULT 14
US-09-513-996A-50583
; Sequence 50583, Application US/09513996A
; GENERAL INFORMATION:
; APPLICANT: N. ALEXANDROV et al.
; TITLE OF INVENTION: SEQUENCE-DETERMINED DNA FRAGMENTS AND CORRESPONDING POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 2750-709P
; CURRENT APPLICATION NUMBER: US/09/513, 996A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 81028
; SEQ ID NO 50583
; LENGTH: 1444
; TYPE: DNA
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 1..1444
; OTHER INFORMATION: any n or Xaa = unknown
; FEATURE:
; OTHER INFORMATION: Location 1..1444 / Ceres Seq. ID 2065831
US-09-513-996A-50583

```

1
2
3
4

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:39:27 : Search time 146 Seconds
(without alignments)
2287.313 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccggcgccggcgtccaccagct.....ggcatacttctgtgtcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 655294 seqs, 415357811 residues

Total number of hits satisfying chosen parameters: 1310588

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_MA_New:*
1: /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	88.8	22.1	533	US-09-531-113-41876	Sequence 41876, A
2	88	21.9	1184	PCT-US02-28445-13	Sequence 13, Appl
3	88	21.3	539	US-10-236-699-13	Sequence 13, Appl
4	85.6	21.3	539	US-09-531-113-2951	Sequence 2951, Ap
5	61.2	15.2	1014	PCT-US02-28445-5	Sequence 5, Appl1
6	61.2	15.2	1014	US-10-236-699-5	Sequence 5, Appl1
7	33.8	8.4	183	US-10-092-411A-981	Sequence 981, App
8	33.8	8.4	1291	US-09-724-676-16911	Sequence 16911, A
9	33.8	8.4	1291	US-09-724-676A-16911	Sequence 16911, A
10	33.8	8.4	1317	US-09-724-676A-16914	Sequence 16914, A
11	33.8	8.4	1317	US-09-724-676A-16914	Sequence 16914, A
12	33.8	8.4	1358	US-09-724-676-16912	Sequence 16912, A
13	33.8	8.4	1358	US-09-724-676A-16912	Sequence 16912, A
14	33.8	8.4	1384	US-09-724-676-16915	Sequence 16915, A
15	33.8	8.4	1384	US-09-724-676A-16915	Sequence 16915, A
16	33.8	8.4	1388	US-09-724-676-16910	Sequence 16910, A
17	33.8	8.4	1388	US-09-724-676A-16910	Sequence 16910, A
18	33.8	8.4	1414	US-09-724-676-16913	Sequence 16913, A
19	33.8	8.4	1414	US-09-724-676A-16913	Sequence 16913, A
20	33.8	8.4	1544	US-09-724-676-16905	Sequence 16905, A
21	33.8	8.4	1544	US-09-724-676A-16905	Sequence 16905, A
22	33.8	8.4	1570	US-09-724-676-16908	Sequence 16908, A
23	33.8	8.4	1570	US-09-724-676A-16908	Sequence 16908, A
24	33.8	8.4	1611	US-09-724-676-16906	Sequence 16906, A
25	33.8	8.4	1611	US-09-724-676A-16906	Sequence 16906, A
26	33.8	8.4	1637	US-09-724-676-16909	Sequence 16909, A

27	33.8	8.4	1637	US-09-724-676A-16909	Sequence 16909, A
28	33.8	8.4	1641	US-09-724-676-16904	Sequence 16904, A
29	33.8	8.4	1641	US-09-724-676A-16904	Sequence 16904, A
30	33.8	8.4	1667	US-09-724-676-16907	Sequence 16907, A
31	33.8	8.4	1667	US-09-724-676A-16907	Sequence 16907, A
32	33.8	8.4	2172	US-09-724-676-16919	Sequence 16919, A
33	33.8	8.4	2172	US-09-724-676A-16919	Sequence 16919, A
34	33.8	8.4	2211	US-09-724-676-16921	Sequence 16921, A
35	33.8	8.4	2211	US-09-724-676A-16921	Sequence 16921, A
36	33.8	8.4	2213	US-09-724-676-16917	Sequence 16917, A
37	33.8	8.4	2213	US-09-724-676A-16917	Sequence 16917, A
38	33.8	8.4	2237	US-09-724-676-16902	Sequence 16902, A
39	33.8	8.4	2237	US-09-724-676A-16902	Sequence 16902, A
40	33.8	8.4	2239	US-09-724-676-16920	Sequence 16920, A
41	33.8	8.4	2239	US-09-724-676A-16920	Sequence 16920, A
42	33.8	8.4	2269	US-09-724-676-16918	Sequence 16918, A
43	33.8	8.4	2269	US-09-724-676A-16918	Sequence 16918, A
44	33.8	8.4	2278	US-09-724-676-16900	Sequence 16900, A
45	33.8	8.4	2278	US-09-724-676A-16900	Sequence 16900, A

ALIGNMENTS

```

RESULT 1
US-09-531-113-41876/c
: Sequence 41876, Application US/09531113
: GENERAL INFORMATION:
: APPLICANT: Byrum, Joseph R.
: APPLICANT: Heck, Gregory R.
: APPLICANT: La Rosa, Thomas J.
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Plants
: FILE REFERENCE: 38-21(15761)B
: CURRENT FILING DATE: 2000-03-22
: NUMBER OF SEQ ID NOS: 48629
: SEQ ID NO 41876
: LENGTH: 533
: TYPE: DNA
: ORGANISM: Glycine max
: OTHER INFORMATION: Clone ID: jC-gmrd02910026c05a1
US-09-531-113-41876

Query Match          22.1%   Score 88.8:   DB 5:   Length 533;
Best Local Similarity 69.8%:   Pred. No. 4.2e-17;
Matches 162; Conservative 0; Mismatches 67; Indels 3; Gaps 3;

OY 5 GCCGGCTCACACAGCTGCGCCAGTGAAGGCTTAAGACATGCGCAGAGAGGGCCTT 64
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 331 GTGGGGCACACACTGCTTCAAGAAGTCTTAGTATGTTCCAGATPA-AAGCCTT 273

OY 65 GAACCTGTGTGCTGTGCACCCCTAATTAATGCTCAGATGTGCAATGAGCTTATATTA 124
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 272 GTACTGTATGTGCTCGCACCTAATTACTGTACCGTGTGGAATGTAGCTTATCTT 213

OY 125 AGCTTCATGAGAGAGAGAGAGATGTCAAGCTTTCACAGACCCAGAGAAACAA 184
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 212 AGTTTCATGAGAAATATGAGAA-AAGAGTTAAATTTTTCATGAAACAGAGAGAACAA 154

OY 185 CCAGATTCACAGAGGCAAGGCTTGAATCCCATATTTCCCTTTCAGCTCAGCT 236
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 CCAATGAGAGAGGCCCCAGAGAC-AGGCGTCCCATATTTCTTATTAATTGATGT 103

RESULT 2
PCT-US02-28445-13
: Sequence 13, Application PC/TUS0228445
: GENERAL INFORMATION:
: APPLICANT: BASF PLANT SCIENCE GMBH
: TITLE OF INVENTION: PROTEIN PHOSPHATASE STRESS-RELATED POLYPEPTIDES AND
: TITLE OF INVENTION: METHODS OF USE IN PLANTS
: FILE REFERENCE: 16313-0160

```


LENGTH: 1014
TYPE: DNA
ORGANISM: Physcomitrella patens
PCT-US02-28445-5

Query Match 15.2%; Score 61.2; DB 1; Length 1014;
Best Local Similarity 61.9%; Pred. No. 9.6e-09;
Matches 130; Conservative 0; Mismatches 78; Indels 2; Gaps 2;

QY 5 GCCGGGCTCACACCTGCGCCAGGAGGCTTAAGACATGTGCGCAGAGAGGCGCTT 64
DB 730 GCCGTGGCATCATTTGCTCAAGAGGAGTCACTGCTTCTGACAA-AGGACTT 788
QY 65 GAACCTGTGTGCTGCTGACCTTAATTTGCTCAGATGTGCAATGGAGCTTCTATATTA 124
DB 789 GTACAGGTGTGCTGCTGCTCAAACTATGCTACAGATGTGGAATGTGCTTCAATCTTA 848
QY 125 AGCTTCATGAGAGAGAGAGAGATGTCAAGTCTTCACAGACCGAGAGAAACAA 184
DB 849 AGCTTCAACGAAATATGAGAGAGATGTGAATTTTCTACAGACCGAGAGAAACAA 907
QY 185 CCAGATTCAGAGGCGGCTTGGAATCC 214
DB 908 GGCTATGATGGCACCTCGACGAGGATTC 937

RESULT 6
US-10-236-699-5
Sequence 5, Application US/10236699

GENERAL INFORMATION:
APPLICANT: BOHNERT, HANS J.
APPLICANT: CHEN, RUOYING
APPLICANT: ISHITANI, MANABU
APPLICANT: VAN THIELEN, NOCHIA
APPLICANT: COSTA E SILVA, OSWALDO DA
TITLE OF INVENTION: PROTEIN PHOSPHATASE STRESS-RELATED POLYPEPTIDES AND
FILE REFERENCE: 16313-0161
CURRENT APPLICATION NUMBER: US/10/236,699
CURRENT FILING DATE: 2002-09-05
PRIOR APPLICATION NUMBER: 60/317,305
PRIOR FILING DATE: 2001-09-05
NUMBER OF SEQ ID NOS: 43
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1014
TYPE: DNA
ORGANISM: Physcomitrella patens
US-10-236-699-5

Query Match 15.2%; Score 61.2; DB 6; Length 1014;
Best Local Similarity 61.9%; Pred. No. 9.6e-09;
Matches 130; Conservative 0; Mismatches 78; Indels 2; Gaps 2;

QY 5 GCCGGGCTCACACCTGCGCCAGGAGGCTTAAGACATGTGCGCAGAGAGGCGCTT 64
DB 730 GCCGTGGCATCATTTGCTCAAGAGGAGTCACTGCTTCTGACAA-AGGACTT 788
QY 65 GAACCTGTGTGCTGCTGACCTTAATTTGCTCAGATGTGCAATGGAGCTTCTATATTA 124
DB 789 GTACAGGTGTGCTGCTGCTCAAACTATGCTACAGATGTGGAATGTGCTTCAATCTTA 848
QY 125 AGCTTCATGAGAGAGAGAGAGATGTCAAGTCTTCACAGACCGAGAGAAACAA 184
DB 849 AGCTTCAACGAAATATGAGAGAGATGTGAATTTTCTACAGACCGAGAGAAACAA 907
QY 185 CCAGATTCAGAGGCGGCTTGGAATCC 214
DB 908 GGCTATGATGGCACCTCGACGAGGATTC 937

RESULT 7
US-10-092-411A-981/c

Sequence 981, Application US/10092411A
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
FILE REFERENCE: 032796-101
CURRENT APPLICATION NUMBER: US/10/092,411A
CURRENT FILING DATE: 2002-03-07
PRIOR APPLICATION NUMBER: US 09/134,001
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5676
SEQ ID NO 981
LENGTH: 183
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-10-092-411A-981

Query Match 8.4%; Score 33.8; DB 6; Length 183;
Best Local Similarity 58.4%; Pred. No. 1;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 256 TATTTAATCTGGAAGGGGTAAAGACATCATGATGTTGTTATTTGAGGTTGATC 315
DB 175 TTTTAAATCTGAGAGACTTCGTTATCATGAGATTAATTTTCTTCATCATTTT 116
QY 316 ATAAATGCAATCTGATGATTCATTCATTTCTGTTCTTTGAA 356
DB 115 CCTGTGAATCTGTTCCATTTCTTTCTTAATCTTTTAA 75

RESULT 8

US-09-724-676-16911/c
Sequence 16911, Application US/09724676
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 16911
LENGTH: 1291
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: misc_feature
LOCATION: (31)..(31)
OTHER INFORMATION: n is a,c,g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (34)..(34)
OTHER INFORMATION: n is a,c,g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (41)..(41)
OTHER INFORMATION: n is a,c,g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (46)..(46)
OTHER INFORMATION: n is a,c,g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51)..(51)
OTHER INFORMATION: n is a,c,g, or t
FEATURE:
NAME/KEY: misc_feature
LOCATION: (56)..(56)
OTHER INFORMATION: n is a,c,g, or t

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (60)..(60)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-16911
```

```
Query Match      8.4%; Score 33.8; DB 5; Length 1291;
Best Local Similarity 52.5%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
```

```
OY 254 AATATTAAATCGAGAGGCTAAAGAACATCATGTGTTAATGTGGAGTTCCAGA 313
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1262 AATTTTACGCTGCTTTTGGGTTTATACAAATGAACATGTATTATGTAGAGAAACG 1203
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 314 TCATATGCGCATCTGATGCATTCATTTCTTTGACACAGTACAAACAAAC 373
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1202 ATGTATCTCTTATATATAAATCCATTTCCATTTAGCTTTTAAAAACAAAGCT 1143
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 374 AACGGAGGGGCGATACCTCC 394
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1142 GTTGTGACAGATGAACATCC 1122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 9
US-09-724-676A-16911/c
; Sequence 16911, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724.676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16911
; LENGTH: 1291
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34)..(34)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (41)..(41)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (46)..(46)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51)..(51)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (56)..(56)
; OTHER INFORMATION: n is a,c,g, or t
; NAME/KEY: misc_feature
; LOCATION: (60)..(60)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-16911
```

```
Query Match      8.4%; Score 33.8; DB 5; Length 1291;
Best Local Similarity 52.5%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
```

```
OY 254 AATATTAAATCGAGAGGCTAAAGAACATCATGTGTTAATGTGGAGTTCCAGA 313
```

```
DB 1262 AATTTTACGCTGCTTTTGGGTTTATACAAATGAACATGTATTATGTAGAGAAACG 1203
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 314 TCATATGCGCATCTGATGCATTCATTTCTTTGACACAGTACAAACAAAC 373
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1202 ATGTATCTCTTATATATAAATCCATTTCCATTTAGCTTTTAAAAACAAAGCT 1143
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 374 AACGGAGGGGCGATACCTCC 394
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1142 GTTGTGACAGATGAACATCC 1122
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 10
US-09-724-676-16914/c
; Sequence 16914, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724.676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 16914
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (34)..(34)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (41)..(41)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (46)..(46)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51)..(51)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (56)..(56)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (60)..(60)
; OTHER INFORMATION: n is a,c,g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (182)..(182)
; OTHER INFORMATION: n is a,c,g, or t
; NAME/KEY: misc_feature
; LOCATION: (188)..(188)
; OTHER INFORMATION: n is a,c,g, or t
; NAME/KEY: misc_feature
; LOCATION: (198)..(199)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-16914
```

```
Query Match      8.4%; Score 33.8; DB 5; Length 1317;
Best Local Similarity 52.5%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
```


Db 1209 GTTGTGACAGATGAACATCC 1189

RESULT 13

US-09-724-676A-16912/C
Sequence 16912, Application US/09724676A

GENERAL INFORMATION:

APPLICANT: Compugen LTD

TITLE OF INVENTION: Variants of alternative splicing

FILE REFERENCE: 129181.4 Compugen

CURRENT APPLICATION NUMBER: US/09/724,676A

CURRENT FILING DATE: 2000-11-28

NUMBER OF SEQ ID NOS: 97222

SOFTWARE: PatentIn version 3.2

SEQ ID NO 16912

LENGTH: 1358

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (31)..(31)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (34)..(34)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (41)..(41)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (46)..(46)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (51)..(51)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (56)..(56)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (60)..(60)

OTHER INFORMATION: n is a,c,g, or t

US-09-724-676A-16912

Query Match 8.4%; Score 33.8; DB 5; Length 1358;

Best Local Similarity 52.5%; Pred. No. 1.7;

Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 254 AATATTAAATCTGGAAGGGTAAGAACATCATGTGTTATTTGCGATTTCGA 313

Db 1339 ATTTTATGCTGCTTTTGGGTTTATACAAATGAACATGTATTATTTAGAGAAACG 1270

QY 314 TCATATGCAATCTGATCATTCATTCATTTCTTTTGAACAACAAGTCAAAAAACC 373

Db 1269 ATGTATCTCTTTATGATATAAATCCATTTCCATTTAGCTTTTAAAAAACAAGCT 1210

QY 374 AACGGAGGGGCAATACCTCC 394

Db 1209 GTTGTGACAGATGAACATCC 1189

RESULT 14

US-09-724-676-16915/C

Sequence 16915, Application US/09724676

GENERAL INFORMATION:

APPLICANT: Compugen LTD

TITLE OF INVENTION: Variants of alternative splicing

FILE REFERENCE: 129181.4 Compugen

CURRENT APPLICATION NUMBER: US/09/724,676

CURRENT FILING DATE: 2000-11-28

NUMBER OF SEQ ID NOS: 97222

SOFTWARE: PatentIn version 3.2

SEQ ID NO 16915

LENGTH: 1384

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (31)..(31)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (34)..(34)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (41)..(41)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (46)..(46)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (51)..(51)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (56)..(56)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (60)..(60)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (182)..(182)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (188)..(188)

OTHER INFORMATION: n is a,c,g, or t

FEATURE:

NAME/KEY: misc_feature

LOCATION: (198)..(199)

OTHER INFORMATION: n is a,c,g, or t

US-09-724-676-16915

Query Match 8.4%; Score 33.8; DB 5; Length 1384;

Best Local Similarity 52.5%; Pred. No. 1.7;

Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 254 AATATTAAATCTGGAAGGGTAAGAACATCATGTGTTATTTGCGATTTCGA 313

Db 1335 ATTTTATGCTGCTTTTGGGTTTATACAAATGAACATGTATTATTTAGAGAAACG 1296

QY 314 TCATATGCAATCTGATCATTCATTCATTTCTTTTGAACAACAAGTCAAAAAACC 373

Db 1225 ATGTATCTCTTTATGATATAAATCCATTTCCATTTAGCTTTTAAAAAACAAGCT 1236

QY 374 AACGGAGGGGCAATACCTCC 394

Db 1235 GTTGTGACAGATGAACATCC 1215

RESULT 15

US-09-724-676A-16915/C

Sequence 16915, Application US/09724676A

GENERAL INFORMATION:

APPLICANT: Compugen LTD

TITLE OF INVENTION: Variants of alternative splicing

FILE REFERENCE: 129181.4 Compugen

```
/ CURRENT APPLICATION NUMBER: US/09/724,676A
/ CURRENT FILING DATE: 2000-11-28
/ NUMBER OF SEQ ID NOS: 97222
/ SOFTWARE: Patentin version 3.2
/ SEQ ID NO 16915
/ LENGTH: 1384
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (31)..(31)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (34)..(34)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (41)..(41)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (46)..(46)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51)..(51)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (56)..(56)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (60)..(60)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (182)..(182)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (188)..(188)
/ OTHER INFORMATION: n is a,c,g, or t
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (198)..(199)
/ OTHER INFORMATION: n is a,c,g, or t
/ US-09-724-676A-16915

Query Match      8.4%; Score 33.8; DB 5; Length 1384;
Best Local Similarity 52.5%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 254 AATATTAAATCGGAAGGGGTAACAATGAGATGCGTAATTGTCGAGTTCAGA 313
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1355 ATTTTACGTGCTTTTGGGTTTATACATGACATGATTAATGTAAGAAAAACG 1296

QY 314 TCATAATGCCATCTGATGATTCATTTCTGTTCTTTGAACACAGTCAAAAAACC 373
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1295 ATGTCATCCCTTTATGATTAATTCATTTTCATTTTAGCTTTTAAAAAACAAGCT 1236

QY 374 AACGGAGGGGCAATTAACCTCC 394
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1235 GTTGTGACAGATGAACATCC 1215
```

Search completed: January 2, 2003, 15:04:23
Job time : 149 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 2, 2003, 12:30:02 : Search time 2188 Seconds
(without alignments)
2975.588 Million cell updates/sec

Title: US-09-850-147-1
Perfect score: 402
Sequence: 1 ccggcgcgcgcgcctaccagctc.....ggcatactctcgtgttcc 402

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 100%
Listing first 45 summaries

Database :
EST: *
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	266	66.2	568	10	AM565695 LG1_348_A
2	266	66.2	606	14	BQ656076 PR0407 S
3	266	66.2	612	10	AM745589 WSL_35_A0
4	266	66.2	614	10	BE597103 P11_70_GO
5	266	66.2	681	13	BG946826 IPI_1.B10
6	252.4	62.8	493	12	BG558648 RH1Z2_58_

Result No.	Score	Query Match	Length	ID	Description
7	229.4	57.1	479	10	AM091358
8	229.4	57.1	570	10	AT978092
9	229.4	57.1	592	10	AM036911
10	227.8	56.7	1349	11	AT109457
11	221.8	55.2	470	9	AT621443
12	220	54.7	590	10	AM288508
13	220	54.7	613	9	AT691590
14	218	54.2	413	10	AM924010
15	212.4	52.8	605	13	BM079618
16	210.4	52.3	537	12	BG842699
17	195	48.5	365	13	BI388689
18	195	48.5	400	13	BI396083
19	186.4	46.4	352	12	BG354266
20	181	45.0	487	10	BE598350
21	178.6	44.4	604	13	BQ284678
22	175.8	43.7	301	9	AT902083
23	173.8	43.2	634	13	BQ251011
24	167.4	41.6	535	10	AM679158
25	166.6	41.4	414	9	AU031687
26	154.8	38.5	570	13	BQ466432
27	153.2	38.1	356	9	AT619388
28	152.6	38.0	492	10	AV943812
29	152.6	38.0	653	13	BQ482869
30	152.4	37.9	665	13	BQ211965
31	151	37.6	536	9	AT435241
32	151	37.6	547	14	BQ664779
33	149.2	37.1	578	10	AM065323
34	148.8	37.0	511	9	AT979736
35	148.6	37.0	593	9	AL499918
36	140.4	34.9	587	13	BQ486368
37	138.4	34.4	531	12	BG265579
38	116	28.9	693	13	BQ322756
39	112.6	28.0	630	14	BQ472249
40	111	27.6	707	10	AV946097
41	108.8	27.1	610	9	AL500634
42	107.2	26.7	477	12	BF199841
43	93.6	23.3	624	14	BQ510007
44	93	23.1	539	10	AM041153
45	92	22.9	481	14	BQ120114

ALIGNMENTS

RESULT 1
AM565695
LOCUS
DEFINITION LG1_348_A05.g1_A002 light grown 1 (LG1) Sorghum bicolor CDNA, mRNA
ACCESSION AM565695
VERSION AM565695.1 GI:7219573
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 568)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
TITLE An EST database from Sorghum: light-grown seedlings
JOURNML Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmprratt@uga.edu
Sequences have been trimmed to exclude polyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: T7
High quality sequence start: 29

High quality sequence stop: 568
POLYA=Yes.

FEATURES	Location/Qualifiers
source	1. .568

BASE COUNT	BRIGIN
147 a	99 c 142 g 180 t

contig	Seq primer: M13F/M13R.
Location/Qualifiers	
1. .606	

```

/organism="Sorghum halepense"
/db_xref="taxon:4560"
/clone="PRC0407"
/clone_lib="Sorghum halepense rhizome cDNA library"
/tissue_type="rhizome"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI; cDNA library was prepared from mRNA isolated from the apical 2-3 cm of actively growing rhizomes of Johnson grass (Sorghum halepense) and cloned into the EcoRI/XhoI sites of lambda ZAP II (Stratagene)."
BASE COUNT      152 a      110 c      147 g      197 t
ORIGIN

```

RESULT 2	
LOCUS	B0656076
DEFINITION	B0656076 606 bp mRNA linear EST 15-JUL-2002 PBC04.07 Sorghum halepense rhizome cDNA library Sorghum halepense
ACCESSION	CNNA clone PRC0407, mRNA sequence.
VERSION	B0656076
KEYWORDS	B0656076.1 GI:21788402 EST.
SOURCE	Sorghum halepense. Sorghum halepense
ORGANISM	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE	1 (bases 1 to 606)
AUTHORS	Paterson,A.H.
TITLE	Unpublished, Paterson,A.H
JOURNAL	Unpublished (2002)
COMMENT	Contact: Paterson AH

TITLE	JOURNAL	COMMENT
An EST database from Sorghum: water-stressed plants	Unpublished (2000)	Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860

Fax: 706 583 0210
Email: impratte@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: T7
High quality sequence start: 7
High quality sequence stop: 612
POLYA=Yes.

FEATURES

source

1. .612
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_11b="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."
BASE COUNT 156 a 110 c 148 g 198 t
ORIGIN

Query Match 66.2%; Score 266; DB 10; Length 612;
Best Local Similarity 84.2%; Pred. No. 6.3e-71;
Matches 335; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

OY 5 GCCGGGCTCACACCTGCGCCGAGGAGGCTTAAGACATGTGCGAGAGAGGGCCTT 64
DB 97 GCCGGGCTCACACCTGCGCCGAGGAGGCTTAAGACATGTGCGAGAGAA-GGGCCTT 155
OY 65 GAACCTGTGTGCTGACCCCTAATTTATGCTGAGATGTGCAATGAGCTTCTATATTA 124
DB 156 GTACCTGTGTGCTGACCCCTAATTTATGCTGAGATGTGCAATGAGCTTCTATATTA 215
OY 125 AGCTTCGATGAGAAGAAAGAGATGTCAGGCTTTCACAGACCGAGAAACAA 184
DB 216 AGCTTCGATGAGAAGATGAGAA-GAGATGTCAGGCTTTCACAGACAGAGAGAAACAA 274
OY 185 CCAGATCCAGGGGCGGAGGCTTGAATCCCATTTCTTGTGACTGACGTGACATTT 244
DB 275 CCAGATCCAGGGGCGGAGGAGGAC-TGCAGTCCCATTTCTGCTGACGTGACATTT 333
OY 245 TTAGAAGGCATATTTAAATCTGGAAGGGGTAAGAACAATCAATGAGTGTAAATTTG 304
DB 334 TTAGAAGGCATATTTAAATCTGTAAGGGTAAAGATCAATCAATGAGTGTAAATTTG 393
OY 305 AGGTTGATCATATATGCAATCTGATGATTCATTCATTTCTGTTCTTTTGAACAAGTA 364
DB 394 AGGTTGATCATATGATGCAATCTGATGATTCATTCATTTCTGTTCTTTTGAACAAGTA 453
OY 365 CAAAAAACCAAGGAGGAGGATTAATCTGCTGCTTCC 402
DB 454 CAATGACCATCGGAGGTGATCATTTCTGTAATTC 491

RESULT 4
BE597103 614 bp mRNA linear EST 18-AUG-2000
LOCUS P11_70_G02_g1_A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION BE597103
VERSION BE597103.1 GI:9852288
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 614)
Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL unpublished (2000)

COMMENT

Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: impratte@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTMix
High quality sequence start: 12
High quality sequence stop: 604
POLYA=No.

FEATURES

source

1. .614
Location/Qualifiers
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_11b="Pathogen induced 1 (P11)"
/note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation; Vector: pluescript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (Brix 623 cultivar) were infected with pathogen (isolate FRM421 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings 48 hours after inoculation. Note: young seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."
BASE COUNT 156 a 110 c 154 g 193 t 1 others
ORIGIN

Query Match 66.2%; Score 266; DB 10; Length 614;
Best Local Similarity 84.2%; Pred. No. 6.4e-71;
Matches 335; Conservative 0; Mismatches 60; Indels 3; Gaps 3;

OY 5 GCCGGGCTCACACCTGCGCCGAGGAGGCTTAAGACATGTGCGAGAGAGGGCCTT 64
DB 123 GCCGGGCTCACACCTGCGCCGAGGAGGCTTAAGATGATGTCAGAGAGAA-GGGCCTT 181
OY 65 GAACCTGTGTGCTGACCCCTAATTTATGCTGAGATGTGCAATGAGCTTCTATATTA 124
DB 182 GTACCTGTGTGCTGACCCCTAATTTATGCTGAGATGTGCAATGAGCTTCTATATTA 241
OY 125 AGCTTCGATGAGAAGAAAGAGATGTCAGGCTTTCACAGACCGAGAAACAA 184
DB 242 AGCTTCGATGAGAAGATGAGAA-GAGATGTCAGGCTTTCACAGACAGAGAGAAACAA 300
OY 185 CCAGATCCAGGGGCGGAGGCTTGAATCCCATTTCTTGTGACTGACGTGACATTT 244
DB 301 CCAGATCCAGGGGCGGAGGAGGAC-TGCAGTCCCATTTCTGCTGACGTGACATTT 359
OY 245 TTAGAAGGCATATTTAAATCTGGAAGGGGTAAGAACAATCAATGAGTGTAAATTTG 304
DB 360 TTAGAAGGCATATTTAAATCTGTAAGGGTAAAGATCAATCAATGAGTGTAAATTTG 419
OY 305 AGGTTGATCATATATGCAATCTGATGATTCATTCATTTCTGTTCTTTTGAACAAGTA 364
DB 420 AGGTTGATCATATGATGCAATCTGATGATTCATTCATTTCTGTTCTTTTGAACAAGTA 479
OY 365 CAAAAAACCAAGGAGGAGGATTAATCTGCTGCTTCC 402
DB 480 CAATGACCATCGGAGGTGATCATTTCTGTAATTC 517

RESULT 5

BC946826	681 bp	mRNA	linear	EST 12-JUN-2001					
LOCUS	BC946826	IP1_1.B10_g1.A002	Immature panicle 1 (IP1)	Sorghum bicolor CDNA,					
DEFINITION	mRNA sequence.								
ACCESSION	BC946826								
VERSION	BC946826.1	GI:14365015							
KEYWORDS	EST.								
SOURCE	Sorghum.								
ORGANISM	Sorghum bicolor								
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Sorghum.								
AUTHORS	Klein,R.R., Cordomier-Pratt,M.-M., Gingle,A., Sudman,M. and Pratt ,L.H.								
TITLE	An EST database from Sorghum: developing preanthesis panicles								
JOURNAL	Unpublished (2001)								
COMMENT	Contact: Cordomier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement. Seq primer: T7 High quality sequence start: 56 High quality sequence stop: 681 PolyA-ies.								
FEATURES	Location/Qualifiers								
source	1..681 /organism="Sorghum bicolor" /cultivar="Brx623" /db_xref="taxon:4558" /clone-lib="Immature panicle 1 (IP1)" /note="Organ: Developing preanthesis panicles; Vector: pBluescript II SK(-) from lambda zap II; Site.1: XhoI; Site.2: EcoRI; The library was made from Poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."								
BASE COUNT	177 a 129 c 163 g 212 t								
ORIGIN									
Query Match	66.2%; Score 266; DB 13; Length 681;								
Best Local Similarity	84.2%; Pred. No. 6.5e-71;								
Matches	335; Conservative 0; Mismatches 60; Indels 3; Gaps 3;								
QY	5	GCCGGGCTCACGACGTGCGCCAGTGAAGGCTTAAAGACATGTGCCAGAGAGAAGGCGCTT	64						
DB	255	GCCGGGCTCACGACGTGCTCCAGAAAGCTTAAATACATGTTCCAGAGAA-GGGCTT	313						
QY	65	GAACCTGTGTGTCTGCACCCCTAATTATTTGCTCAATGTGGCAATGGAGCTTCTATATTA	124						
DB	314	GTAACCTGTGTGTCTGCACCTAATTATTTGCTCAACATGTGGCAATGTAGCTTCTATATTA	373						
QY	125	AGCTTCGATGGAAGAAAGAAAGGATGTAAGGTCTTCACAGAGACCGAGGAACAA	184						
DB	374	AGCTTCGATGGAAGAAAGGATGTAAGGTCTTCACAGAGACCGAGGAACAA	432						
QY	185	CCAGATTCAGAGGCGCAAGGCTTGAATCCCATATTTCTTGAATCAGGTGCAATTT	244						
DB	433	CCAGATTCAGAGGCGCAAGGCTTGAATCCCATATTTCTTGAATCAGGTGCAATTT	491						
QY	245	TTGAGAGCAATATTTAAATCTGCAAGGGGTAAAGAACATTCAGATGCTTAAATTGGG	304						
DB	492	TTGAGAGCAATATTTAAATCTGCAAGGGGTAAAGAACATTCAGATGCTTAAATTGGG	551						
QY	305	AGGTTCACATCATATATGCATCTGATGCATCCATTTCTGTCTTTTGAACACAACTA	364						
DB	552	AGGTTCACATCATATATGCATCTGATGCATCCATTTCTGTCTTTTGAACACAACTA	611						

QY	365	CAAAAAACCAAGGAGGGGCATCTTCGTGTCC	402
Dd	612	CAATAGACCATCGGAGTGTCTATCTTCCTATTCC	649
RESULT 6			
LOCUS	BG558648		
DEFINITION	BG558648 RH1Z2.58_E10_g1_A003 Rhizome2 (RH1Z2) Sorghum propinquum cDNA, mRNA sequence.	493 bp	EST 10-APR-2001
ACCESSION	BG558648		
VERSION	BG558648.1	GI:13587646	
KEYWORDS	EST.		
SOURCE	Sorghum propinquum.		
ORGANISM	Sorghum propinquum Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.		
REFERENCE	1 (bases 1 to 493)		
AUTHORS	Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt,L.H.		
TITLE	An EST database from Sorghum: Sorghum propinquum rhizomes unpublished (2000)		
JOURNAL	Contact: Cordonnier-Pratt MM		
COMMENT	Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Km. 2502, Athens, GA 30602-7271, USA Tel.: 706 542 1860 Fax: 706 583 0210 Email: mmprratl@uga.edu Seq primer: T7 High quality sequence start: 70 High quality sequence stop: 419 POLYA-Tes. Location/Qualifiers 1..493 /organism="Sorghum propinquum" /db_xref="taxon:132711" /clone_lib="Rhizome2 (RH1Z2)" /note="Organ: Rhizomes; Vector: pluscript II from Lambda Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision."		
BASE COUNT	131 a 90 c 115 g 157 t		
ORIGIN			
Query Match	62.8%; Score 252.4; DB 12; Length 493;		
Best Local Similarity	83.7%; Pred. No. 9.4e-67;		
Matches 333; Conservative	0; Mismatches 61; Indels 4; Gaps 4;		
QY	5	GCCGGGCTCACAGTGCCGCCAAGAAGCTTAAGCATGTGCCAGGAAGGGCCCTT	64
Dd	50	GCCGGGCTCACAGTGCGGCCAGAAAGCTTTAAATCAATGTTCCAGAGAA-GGGCCTT	108
QY	65	GAACTTGTGTGTCTGCACCCCTAATTATTGTGCAGATGTGGCAATGAGCTTCTATTA	124
Dd	109	GTAACCTGTGTGTCTGCACCTAATTATTGTGCACATGTGGGAATGTAGCTTCTATTA	168
QY	125	AGCTTCGATGAGAGAAAGGATGTCAAAGTCTTACACAGACCCAGAGAAAACA	184
Dd	169	AGCTTCGATGAGAAAGATGAAA-GAGATGTCAAGTCTTACAGACAGAGAGAAACA	227
QY	185	CCAGATCCAGGGCCCAAGGGCTTGAATCCCATATTCTTGGACTTCAGGTGGACATT	244
Dd	228	CCAGATGGAGGCCCAAGGAC-TGAGTCCCATATTCTCTGACTTGAAGTGTACAAATT	286
QY	245	TTAGAGCCAAATTATTAATCTGGAAGGGTAAAGAAATCAGATGTGTAAATTGGG	304
Dd	287	TTAGAGTCAANTATTAATCTGTAAAGGTTAAAGATCAATCAGATGTGTATTGTCT	346
QY	305	AGGTTCAATCAATAATGCCATCTGATGCATTCATTCCTGTGTTTTTGAACACAACTA	364

Db 347 AGTTCAGATCATCTAGTCATTCGATTCGATTCCTGTTCTTTTGACCATGTA 406
Oy 365 CAAAACCAACGAGGAGGAGGATTAAGTTCCTGTTGCC 402
Db 407 CAATGACCTGCGGAGGTGTCA-ATCTTCTCTGTTATTC 443

RESULT 7
AM091358/c 479 bp mRNA linear EST 18-OCT-1999
LOCUS 614094A10.x1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION AM091358
VERSION AM091358.1 GI:6056953
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 479)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614094 row: A column: 10.

FEATURES
source
1..479
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/rissue_type="root"
/dev_stage="3-4 days old"
/lab_host="X10LR"
/note="Organ: root; Vector: pBluescriptII SK+; Site_1:
EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
Lab (LM)";
BASE COUNT 142 a 111 c 96 g 130 t
ORIGIN

Query Match 57.1%; Score 229.4; DB 10; Length 479;
Best Local Similarity 81.7%; Pred. No. 1.1e-59;
Matches 313; Conservative 0; Mismatches 66; Indels 4; Gaps 4;

Oy 5 GCCGGGCTCACCAGCTGGCCAGTGAAGGCTTAAGACATGTGCCAGAGAGAGGGCCTT 64
Db 405 GCCGGGCTCACCAGCTGGCTGCAGAGAGGCTTAAGTACATGTTCACAGAGAA-GGGCCTT 347
Oy 65 GAACCTGTGTGTGCTGCACCCCTAATTATGCTCAGATGTGGCAATGAGCTTCTATATA 124
Db 346 GTAACGTGTGTGTGCTGCACCCCAATTATGCTACAGATGTGGCAATGTAGCTTCTATATA 287
Oy 125 AGCTTCGATGAGAAAGAAAGAGATGTCAAGCTCTTCACAGAGACCGAGAAACAA 184
Db 286 AGCTTCGATGAGAAAGATGAAA-GAGATGTCAAGTCTTCACAGAGAGAGAAACAA 228
Oy 185 CCAGATTCAGGGCCAGAGGGCTTGAATCCATATTTCTTGAAGTGAAGTGAATTT 244
Db 227 CCAGATTCAGAGGCCCAAGAGAC-TGCAGTCCCATATTTCTCTGACTTGAATTTGTAC-ATT 170
Oy 245 TTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAAATGATGCTTAATTTGG 304
Db 169 TTAGAGTCAACATTTAAATCTGTAAGGGCTTAAGATCAGTCAGATGTGTATTCATTGT 110
Oy 305 AGTTCAGATCATTAATGCAATCTGATGATTCATTTCTGTTCTTTGAACACAGTA 364

Db 109 AGTTCAGATCATCTAGTCATTCGATTCGATTCCTGTTCTTTTAACACATGTA 50
Oy 365 CAAAACCAACGAGGAGGAGGATTAAGTTCCTGTTGCC 387
Db 49 CAATGACCTGCGGAGGTGTCA-ATCTTCTCTGTTATTC 27

RESULT 8
A1978092/c 570 bp mRNA linear EST 27-AUG-1999
LOCUS 614022D04.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
DEFINITION mRNA sequence.
ACCESSION A1978092
VERSION A1978092.1 GI:5791300
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 570)
AUTHORS Walbot, V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 614022 row: D column: 04.

FEATURES
source
1..570
/organism="Zea mays"
/cultivar="W23"
/db_xref="taxon:4577"
/clone_lib="614 - root cDNA library from Walbot Lab"
/rissue_type="root"
/dev_stage="3-4 days old"
/lab_host="X10LR"
/note="Organ: root; Vector: pBluescriptII SK+; Site_1:
EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
Lab (LM)";
BASE COUNT 177 a 133 c 112 g 148 t
ORIGIN

Query Match 57.1%; Score 229.4; DB 9; Length 570;
Best Local Similarity 81.7%; Pred. No. 1.2e-59;
Matches 313; Conservative 0; Mismatches 66; Indels 4; Gaps 4;

Oy 5 GCCGGGCTCACCAGCTGGCCAGTGAAGGCTTAAGACATGTGCCAGAGAGAGGGCCTT 64
Db 497 GCCGGGCTCACCAGCTGGCTGCAGAGAGGCTTAAGTACATGTTCACAGAGAA-GGGCCTT 439
Oy 65 GAACCTGTGTGTGCTGCACCCCTAATTATGCTCAGATGTGGCAATGAGCTTCTATATA 124
Db 438 GTAACGTGTGTGTGCTGCACCCCAATTATGCTACAGATGTGGCAATGTAGCTTCTATATA 379
Oy 125 AGCTTCGATGAGAAAGAAAGAGATGTCAAGCTCTTCACAGAGACCGAGAAACAA 184
Db 378 AGCTTCGATGAGAAAGATGAAA-GAGATGTCAAGTCTTCACAGAGAGAGAAACAA 320
Oy 185 CCAGATTCAGGGCCAGAGGGCTTGAATCCATATTTCTTGAAGTGAAGTGAATTT 244
Db 319 CCAGATTCAGAGGCCCAAGAGAC-TGCAGTCCCATATTTCTCTGACTTGAATTTGTAC-ATT 262
Oy 245 TTAGAGGCAATATTTAAATCTGGAAGGGGTAAAGAACAAATGATGCTTAATTTGG 304
Db 261 TTAGAGTCAACATTTAAATCTGTAAGGGCTTAAGATCAGTCAGATGTGTATTCATTGT 202

QY 305 AGGTCAGATCATTAATGCGCATTCGATTCATTTCTGTTCTTTGAAACAAGTA 364
 |||||
 DB 201 AGGTCAGATCATTAATGCGCATTCGATTCATTTCTGTTCTTTTAAACAAGTA 142
 QY 365 CAAAAAACCAACGGAGGGGCAT 387
 |||||
 DB 141 CAAATAGACCATCCGAGTCTCAT 119

RESULT 9

AM036911 592 bp mRNA linear EST 15-SEP-1999
 LOCUS 614022D04.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
 DEFINITION mRNA sequence.
 AM036911
 ACCESSION AM036911.1 GI:5895665
 VERSION
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 592)

REFERENCE Walbot, V.
 AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford
 TITLE University
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 614022 row: D column: 04.

FEATURES

source 1..592
 Location/Qualifiers
 /organism="Zea mays"
 /cultivar="W23"
 /db_xref="taxon:4577"
 /clone_lib="614 - root cDNA library from Walbot Lab"
 /tissue_type="root"
 /dev_stage="3-4 days old"
 /lab_host="XILOR"
 /note="Organ: root; Vector: pBluescriptII SK+; Site: 1;
 EcorI; Site: 2; XhoI; 3-4 days old root tissue from Walbot
 Lab (LM)"
 BASE COUNT 159 a 120 c 142 g 171 t
 ORIGIN

Query Match 57.1%; Score 229.4; DB 10; Length 592;
 Best Local Similarity 81.7%; Pred. No. 1.2e-59;
 Matches 313; Conservative 0; Mismatches 66; Indels 4; Gaps 4;
 QY 5 GCCGGGCTCACACAGCTGCGCCAGTGAAGCTTAAAGACATGTGCCAGAGAGGGCCCTT 64
 |||||
 DB 199 GCCGGGCTCACACAGCTGCGCCAGTGAAGCTTAAAGACATGTGCCAGAGAGGGCCCTT 257
 QY 65 GAACCTGTGTGCTGCTGACCTTAATTAATGTCAGATGTGGCAATGGAGCTTTATATTA 124
 |||||
 DB 258 GTAACCTGTGTGCTGCTGACCTTAATTAATGTCAGATGTGGCAATGGAGCTTTATATTA 317
 QY 125 AGCTTCGATGAGAAGAAAGAGAGATGTCAAGGCTTCACAGAGACCGAGAAACAA 184
 |||||
 DB 318 AGCTTCGATGAGAAGAGTGAAGAA-GAGATGTCAAGTTCCTTCACAGAGACCGAGAAACAA 376
 QY 185 CCAGATTCAGAGGCCAAGGCGTTGAATCCCATATTTCTTGACTTCAGGTGGACAATT 244
 |||||
 DB 377 CCAGATTCAGAGGCCAAGGCGTTGAATCCCATATTTCTTGACTTCAGGTGGACAATT 434
 QY 245 TTAGAAGCAATTTAAATCTGGAAGGGGTAAGACAATCAGATGTGTTAATTTGTGG 304
 |||||
 DB 435 TTAGAAGCAATTTAAATCTGTAAGGGCTAAAGATCAGATGTGTTAATCATTTGT 494

QY 305 AGGTCAGATCATTAATGCGCATTCGATTCATTTCTGTTCTTTGAAACAAGTA 364
 |||||
 DB 495 AGGTCAGATCATTAATGCGCATTCGATTCATTTCTGTTCTTTTAAACAAGTA 554
 QY 365 CAAAAAACCAACGGAGGGGCAT 387
 |||||
 DB 555 CAAATAGACCATCCGAGTCTCAT 577

RESULT 10

AY109457/c 1349 bp mRNA linear HTC 25-MAY-2002
 LOCUS AY109457/c
 DEFINITION Zea mays C112356_1 mRNA sequence.
 AY109457
 ACCESSION AY109457
 VERSION AY109457.1 GI:21213184
 KEYWORDS HTC.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoidae; Andropogoneae; Zea.
 1 (bases 1 to 1349)

REFERENCE Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,
 Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
 AUTHORS Maize Mapping Project/DuPont Consensus Sequences for Design of
 TITLE Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 1349)
 AUTHORS Coe, E.C.
 TITLE Direct Submission
 JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA

FEATURES

source 1..1349
 Location/Qualifiers
 /organism="Zea mays"
 /db_xref="MaizeDB:630146"
 /db_xref="taxon:4577"
 /clone="C112356.1"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"
 BASE COUNT 350 a 314 c 286 g 310 t 89 others
 ORIGIN

Query Match 56.7%; Score 227.8; DB 11; Length 1349;
 Best Local Similarity 81.5%; Pred. No. 4.7e-59;
 Matches 312; Conservative 0; Mismatches 67; Indels 4; Gaps 4;
 QY 5 GCCGGGCTCACACAGCTGCGCCAGTGAAGCTTAAAGACATGTGCCAGAGAGGGCCCTT 64
 |||||
 DB 496 GCCGGGCTCACACAGCTGCGCCAGTGAAGCTTAAAGACATGTGCCAGAGAGGGCCCTT 438
 QY 65 GAACCTGTGTGCTGCTGACCTTAATTAATGTCAGATGTGGCAATGGAGCTTTATATTA 124
 |||||
 DB 437 GTAACCTGTGTGCTGCTGACCTTAATTAATGTCAGATGTGGCAATGGAGCTTTATATTA 378
 QY 125 AGCTTCGATGAGAAGAAAGAGAGATGTCAAGGCTTCACAGAGACCGAGAAACAA 184
 |||||
 DB 377 AGCTTCGATGAGAAGAGTGAAGAA-GAGATGTCAAGTTCCTTCACAGAGACCGAGAAACAA 319
 QY 185 CCAGATTCAGAGGCCAAGGCGTTGAATCCCATATTTCTTGACTTCAGGTGGACAATT 244
 |||||
 DB 318 CCAGATTCAGAGGCCAAGGCGTTGAATCCCATATTTCTTGACTTCAGGTGGACAATT 261
 QY 245 TTAGAAGCAATTTAAATCTGGAAGGGGTAAGACAATCAGATGTGTTAATTTGTGG 304
 |||||
 DB 260 TTAGAAGCAATTTAAATCTGTAAGGGCTAAAGATCAGATGTGTTAATCATTTGT 201

OY	305	AGCTTCAGATTCMAAATGCGCATCGATGCATTCCATTTCTGGTTCCTTGACACAAGA	364
Db	200	AGCTTCAGATTCMAAATGCGCATCTTAATGCGCTTGATTTCTGTTCCTTAACCACTGTA	141
OY	365	CAAAAACCMAACGGAGGGGCAT	387
Db	140	CAAATGACCATCCGGAGTCTCAT	118
RESULT 11			
A1621443/C			
LOCUS	A1621443	470 bp	mRNA linear EST 21-Apr-1999
DEFINITION	48609J001.xl 486 - leaf primordia cDNA library from Hake lab Zea		
ACCESSION	A1621443		
VERSION	A1621443.1	GI:4630569	
KEYWORDS	EST.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC		
	clade; Panicoideae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 470)		
AUTHORS	Walbot,V.		
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 486090 row: D column: 01. Location/Qualifiers 1..470 /organism="Zea mays" /cultivar="B73" /db_xref="taxon:4577" /clone_lib="486 - leaf primordia cDNA library from Hake lab" /tissue_type="leaf primordia" /dev_stage="P7-P11 leaf" /lab_host="E.coli XL1-Blue MFR/" /note="Organ: shoot; Vector: Lambda zap; Hake lab cDNA library."		
FEATURES			
source			
BASE COUNT	148 a 109 c 94 g 119 t		
ORIGIN			
Query Match	55.2%	Score 221.8:	DB 9; Length 470;
Best Local Similarity	79.2%;	Pred. No. 2.5e-57;	
Matches 313; Conservative	0;	Mismatches 77;	Indels 5; Gaps 4;
OY	5	GCCGGGCTCACGAGTGGGCCAGTAAGGCTTAAGACATGTGCCAGAGAGAGGCCCTT	64
Db	443	GCCGGGCTCACGAGTGGGCCAGTAAGGCTTAAGACATGTGCCAGAGAGAGGCCCTT	385
OY	65	GAACTTGTGTGTGTGTCACCCCTAATTATTTGTCAGATGTGGCAATGGAAGCTTCTATATTA	124
Db	384	GTCACTGTGTGTGTGTCACCCCTAATTATTTGTCAGATGTGGCAATGGAAGCTTCTATATTA	325
OY	125	AGCTTCGATGAAGAAGAAAGAGAGATCAAGGCTTCACAGAGACCGAGAAACA	184
Db	324	AGCTTCGATGAAGAAGATGAAGAAG-GATGTCAAGTCTTCACAGAGACCGAGAGAACCT	266
OY	185	CCAGATTCGAGGCGCAAGGCTTGAATCCATATTTCCCTTGACTTCAGTGCACAATT	244
Db	265	CCAGATTCGAGGCGCAAGGAC-CGAGATCCCATATTTCTATGACTTCGAATTCACAATT	207
OY	245	TTAGAAGCAATATTTAAATCTGGAAGGGGTAAAGACATAGATGCTTAAATTGTGG	304

DB	206	TTAGAAGTC--TATTCGATCTGTGAAGGGCTTAAGATCAATGATGATGATTTATTTGTTT	149
QY	305	AGGTTGACATCTAATTCATCTGATTCATTCATTTCTGTTCTTTTGAACACAACTA	364
DB	148	AGGTCGACATTTATGTCATCTGATGCGTTTGATTTCTGTTTCCCTTAACCATGTA	89
QY	365	CAAAAACCAACGAGGGGCGATCACTTCCTGCTG	399
DB	88	CAATGACCATCGGACAGTCAATGTTCTCTAT	54
RESULT 12			
LOCUS	AM288508/c	590 bp	mRNA linear EST 16-JAN-2000
DEFINITION	618068C02.x2 618 - Inbred Tassel cDNA Library Zea mays cDNA, mRNA sequence.		
ACCESSION	AM288508		
VERSION	AM288508.1	GI:6695400	
KEYWORDS	EST.		
SOURCE	Zea mays.		
ORGANISM	Zea mays.		
REFERENCE	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.		
AUTHORS	1 (bases 1 to 590)		
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 725 8221 Email: walbot@stanford.edu Plate: 618068 row: C column: 02.		
FEATURES	Location/Qualifiers		
SOURCE	1..590		
	/organism="Zea mays"		
	/cultivar="Ohio43"		
	/db_xref="taxon:4577"		
	/clone_lib="618 - Inbred Tassel cDNA Library"		
	/tissue_type="tassel"		
	/dev_stage="tassel length from 0.1 to 2.5 cm"		
	/lab_host="XLOLR"		
	/note="Organ: tassel; Vector: pAD-GAL4-2.1 (Hybrizap); Inbred tassel library from Schmidt lab"		
BASE COUNT	182 a 138 c 112 g 157 t	1 others	
ORIGIN			
Query Match	54.7%; Score 220; DB 10; Length 590;		
Best Local Similarity	81.8%; Pred. No. 9.6e-57;		
Matches 314; Conservative	0; Mismatches 65; Indels 5; Gaps 5;		
QY	5	GCCGGGCTCACCGAGCTGGCCAGTGAAGGCTTAAAGACATGTGCCAGGAGGAGGCTT	64
DB	436	GCGGGGCTCACCGAGCTGGCTGTCGAGGAAGGCTTAAATACATGTTCCAGGAGAA-GGGCCTT	378
QY	65	GAACTTGTGTGTGTGTCACCCCTAATTAATTTGCTCAAGATGTGGCAATGGAGCTTATATTA	124
DB	377	GTAAGTGTGTGTGTGTCACCCCTAATTAATTTGCTCAAGATGTGGCAATGTAGCTTCTATATTA	318
QY	125	AGCTTCGATGAGAAAGAAAGAGATGTCAAGGCTTCACAGAGACGAGAAACAA	184
DB	317	AGCTTCGATGAGAAATGAAA-GAGATGTCAAGTCTTCACAGAGACGAGAAACAA	259
QY	185	CCAGATTCACAGGCGCAAGGCTTGAATCCATATTTCTTGAAGTTCAGGTGCAAT	244
DB	258	CCAGATTCAGGCGCAAGGAC-TGACAGTCCATATTTCTGACATTTGATGAC-ATT	201
QY	245	TTGAGAGCAATATTTAAATCTGGAAGGGGTAAAGACACATGATGATGCTTAAATTTG	304

Db	200	TTAGAGATCAACATTTAAATCTGTAAAGGCTTAAAGATCAGTCAGATGCTTATATATGCT	141
0y	305	AGGTTCAAGTATCTAATATGCATCTGATGCATTCCTCATTTCT-GGTTTCTTTTGAAACACACT	363
Db	140	AGGTTCAAGTATCTAATATGCATCTGATGCATTCCTCATTTCTTTTCTTTTAAACACACTG	81
0y	364	ACAAAAACCAACGAGGAGGCAT	387
Db	80	ACAAATGACCAATCCGAGTCTCAT	57
RESULT 13			
LOCUS	AI691590/c	613 bp	linear
DEFINITION	606022D02.x1 606 - Ear tissue cDNA library from Schmidt lab Zea		
ACCESSION	AI691590		
KEYWORDS	EST.		
SOURCE	Zea mays.		
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.		
REFERENCE	1 (bases 1 to 613)		
AUTHORS	Walbot,V.		
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford University		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Walbot V Department of Biological Sciences Stanford University 855 California Ave, Palo Alto, CA 94304, USA Tel: 650 723 2227 Fax: 650 723 8221 Email: walbot@stanford.edu Plate: 606022 row: D column: 02.		
FEATURES			
source	1..613 location/Qualifiers /organism="Zea mays" /cultivar="Ohio43" /db_xref="taxon:4577" /clone_id="606 - Ear tissue cDNA library from Schmidt lab" /tissue_type="mixed" /dex_stg="ear length from 0.5 cm - 2.0 cm" /lab_host="XJOLR (Stratagene)" /note="Organ: Immature ear; Vector: pBK-CMV; Site_1: EcoRI Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt lab"		
BASE COUNT	178 a 138 c 114 g 183 t		
ORIGIN			
Query Match	54.7%; Score 220; DB 9; Length 613;		
Best Local Similarity	81.8%; Pred. No. 9.7e-57;		
Matches 314:	Conservative 0; Mismatches 65; Indels 5; Gaps 5;		
0y	5	GGCGGCTCACAGCTGGCCCAAGTGAAGGCTTAAAGATGTCAGAGAGAAAGGCGCTT	64
Db	454	GGCGGCTCACAGCTGGCTGTCAGAGAGGCTTAAATGATGTTCCAGAGAA-GGGCGTT	396
0y	65	GAACCTGTGTGTGTCGACCCCTAATATTATGTCAGATGTGGCAATGAGAGCTTCATATTA	124
Db	395	GTAAGTGTGTGTGTCGACCCCAATATTATGTCAGATGTGGCAATGATGATGCTTCTATATTA	336
0y	125	AGCTTCGATGAGAAAGAAAGAGATGTCAGAGTCTTCACAGAGACCGAGAAACAA	184
Db	335	AGCTTCGATGAGAAAGATGAAA-GAGATGTCAGAGTCTTCACAGAGACAGAGAAACAA	277
0y	185	CCAATTTCCAGGCGCAAGGCTTGAATCCCATATTTCTCTTGACTTGAAGTGGACATTT	244
Db	276	CCAGTGTGAGAGCCCAAGAG-TCGAGTCCCATATTTCTCTTGACTTGAATTTGAC-ATT	219

OY	245	TAGAAAGGCAATTTTAAATCTGTGAAGGGCGTAAGAACAATCAGATGGTGTAATTGGC	304
DB	218	TTAGAGATCACATTTTAAATCTGTGAAGGGCGTAAGAATCATGATGATGTTAATTGT	159
OY	305	AGGTCACATCATTAATGCCATCTGATGCATTCATTTCT-GGTTCTTTTGAACAAGT	363
DB	158	AGGTCACATCATTAATGCCATCTGATGCATTTCTTTTCTTTTAAACACATGT	99
OY	364	ACAAAAAACCAACGGAGGGGCAT	387
DB	98	ACAATAAGACCATCCGAGTCTCAT	75
<hr/>			
RESULT 14			
LOCUS	AM924010/c	413 bp	mRNA EST 19-JUL-2000
DEFINITION	WSL_32.C01.b1_A002 Water-stressed 1 (WSL) Sorghum bicolor cDNA,		
ACCESSION	mRNA sequence.		
VERSION	AM924010		
KEYWORDS	AM924010.1 GI:8089835		
SOURCE	EST.		
ORGANISM	Sorghum.		
PROJECT	Sorghum bicolor		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
AUTHORS	Spermatophytes; Magnoliophyta; Liliopsides; Poales; Poaceae; PACC		
	clade: Panicoideae; Andropogoneae; Sorghum.		
	1 (bases 1 to 413)		
	Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt		
	,L.H.		
TITLE	An EST database from Sorghum: water-stressed plants		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mpratt@uga.edu Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20. Seq primer: JEN REV High quality sequence stop: 405 POLYA-No.		
FEATURES	Location/Qualifiers		
source	1..413		
	/organism="Sorghum bicolor"		
	/db_xref="taxon:4558"		
	/clone_lib="Water-stressed 1 (WSL)"		
	/note="Organ: Mix of 5-week old plants on days 7 & 8 after		
	water was withheld; Vector: lambda zap; Site.1: XhoI;		
	Site.2: EcoRI; The library was made from poly-A RNA in the		
	cloning vector lambda Zap II. Clones to be sequenced were		
	prepared by mass excision."		
BASE COUNT	140 a 89 c 75 g 109 t		
ORIGIN			
Query Match	54.2%	Score 218;	DB 10; Length 413;
Best Local Similarity	86.3%;	Pred. No. 3.6e-56;	
Matches 264;	Conservative	0; Mismatches 40;	Indels 2; Gaps 2;
OY	97	CAGATGTGCAGATGAGCGTCTCATTTTAAGCTTCATGAGAAGAGAGAGAGATGTCA	156
DB	405	CAGATGTGCATGTAGCTTCTATTTAAGCTTCATGAGAAGAGAGAAA-GAATGTCA	347
OY	157	AGGTCITCAGAGAGACCGAGAAAACAACAGATTCCAGGGCCAGGGCTTGAATCCA	216
DB	346	AGTCTTCACAGAGACAGAGAGAAACAACAGATCCAGGCCCAAGGAC-TGCAATCCA	288
OY	217	TATTTCTTTGACTTCAGGTGACGACATTTTAGAAGGCATATTTAAATCTGGAAGGGTA	276
DB	287	TATTTCTTTGACTTCAGGTGACGACATTTTAGAAGGCATATTTAAATCTGGAAGGGTA	228

